

# Specification Theory

The treatment of redundancy in generative phonology

Michael Broe

Doctor of Philosophy  
University of Edinburgh  
1993





THE UNIVERSITY *of* EDINBURGH

Thesis scanned from best copy available:  
may contain faint or blurred text, and / or  
cropped or missing pages.



## Declaration

---

I declare that this thesis has been composed by me and that the research which is reported herein has been conducted by myself unless otherwise indicated.

## Acknowledgements

---

A list of names might be thought by some to provide pretty barren reading, but to the traveller looking back over a long and arduous journey, an unadorned list of names is sufficient to evoke all the exhilarations and fatigues of the trip. Here are the people who shaped my course: some have probably forgotten they ever spoke to me, some probably wished we had spoken a little less often; some are now my friends; one is now my wife. I thank them all for their companionship on the road:

Michael Halliday, Vans Dunlop, John Christie, Jim Hurford, Jim Miller, John Anderson, Ewan Klein, Bob Ladd, Ronnie Cann, Mary Tait, Nigel Fabb, Colin Matheson, Antonio Sanfilippo, Robert Dale, Glyn Morrill, John Coleman, John Local, Diana Archangeli, Doug Pulleyblank, Jo Calder, Steven Bird, Jim Scobbie, Mark Ellison, Steve Isard, Janet Pierrehumbert.

And the band. And the ringers. And Mary, again.

## Abstract

---

This thesis presents a principled method of coding redundancy relations directly into phonological representations; an approach which we term STRUCTURED SPECIFICATION. The approach draws on a number of related ideas from the fields of unification-based lexicons, knowledge representation, and the algebraic theory of lattices.

This move has a number of implications for phonological theory. It resolves the long standing problem of the reciprocal dependency of phonological features; it reveals a connection between the notion of distinctiveness and the elsewhere condition; it results in a lexicon structured as a single lattice, a storage structure which has implications for recognition, acquisition, and lexical access; it extends naturally to the notion of default properties, which also exhibit a lattice structure based on subsumption; it provides a straightforward formalization of the notion archiphoneme, and permits us to distinguish partially specified from redundantly specified segments; analogously, it provides a means of making a principled distinction between unspecified and unmarked segments; it resolves formal and conceptual inconsistencies which result from combining underspecification theory and the theory of feature geometry, with respect to both subclassification and the treatment of monovalent features; it results in a non-derivational approach to specification theory, an approach which makes different empirical predictions from those of current underspecification theory, predictions which we show to be correct. Finally, structured specification is fully compatible with a constraint-based approach to phonology, and thus constitutes the first appropriate theory of specification for this emerging paradigm.

# Contents

---

<b>1.</b>	<b>Well-formedness in phonology and syntax</b>	<b>1</b>
1.1.	Phonological well-formedness	1
1.2.	Syntactic well-formedness	4
1.3.	Unification-based syntax	8
1.4.	Unification-based phonology	8
1.5.	Morpheme structure rules revisited	10
1.6.	Specification theory	12
<b>2.</b>	<b>Constraint-based phonology</b>	<b>15</b>
2.1.	Introduction	15
2.2.	Previous approaches	16
2.3.	Prosodic Analysis	20
2.4.	Allen's approach to Grassmann's Law	29
2.5.	Interaction with Bartholomae's Law	33
2.6.	Conventions or Conditions?	36
<b>3.</b>	<b>Unification-based phonology</b>	<b>40</b>
3.1.	Introduction	40
3.2.	Allen's analysis of Harauti	41
3.3.	Rules, representations and unification	53
3.4.	Modelling prosodies by unification	58
3.5.	Unification grammar and systemic grammar	60
<b>4.</b>	<b>The treatment of redundancy in generative phonology</b>	<b>62</b>
4.1.	Introduction	62
4.2.	The problem of redundant information	63
4.3.	Halle's approach to redundancy	73
4.4.	Stanley's approach to redundancy	86
4.5.	The proposal: structured specification	89
<b>5.</b>	<b>Structured specification</b>	<b>91</b>
5.1.	Introduction	91
5.2.	Structured specification	93
5.3.	Partial Specification	106

5.4.	Default Specification	116
<b>6.</b>	<b>Aspects of structured specification</b>	<b>119</b>
6.1.	Introduction	119
6.2.	Utilizing Lattice-theoretic ideas	119
6.3.	Reciprocal dependencies	129
6.4.	The hierarchy of features and gradient processes	133
<b>7.</b>	<b>Structured specification and feature-geometry</b>	<b>136</b>
7.1.	Subclassification and valency	136
7.2.	First order and second order terms	141
7.3.	Redundancy of articulator nodes	148
7.4.	'Uncharacterized' segments	149
<b>8.</b>	<b>Structured specification and the lexicon</b>	<b>151</b>
8.1.	Introduction	151
8.2.	Sequence structure constraints	152
8.3.	The hierarchical lexicon	159
8.4.	Default features and nonmonotonic inheritance systems	163
8.5.	Combining defaults with redundancy	167
8.6.	The expression of phonotactics	168
8.7.	Comparison with Waterson	169
8.8.	Implications for recognition	172
8.9.	Relativized redundancy	174
<b>9.</b>	<b>Structured specification and underspecification theory</b>	<b>179</b>
9.1.	Introduction	179
9.2.	Contrastive specification	179
9.3.	Radical underspecification	188
9.4.	Structured specification vs contrastive specification	190
9.5.	Problems with underspecification	193
<b>10.</b>	<b>The non-derivational approach to specification</b>	<b>210</b>
10.1.	Introduction	210
10.2.	The precedence of rules affecting distinctive features	212
10.3.	Polysystemic effects	213
10.4.	Against derivational specification	229
	<b>Bibliography</b>	<b>237</b>

# 1. Well-formedness in phonology and syntax

---

## 1.1. Phonological well-formedness

Stanley (1967) represents one of the most rigorous attempts in generative phonology to replace a RULE-BASED account of phonological regularities with an account couched in terms of CONDITIONS — statements of logical implication. The particular issue he addresses is the formulation of redundancy statements in generative phonology, statements which had previously been expressed using rules which were not formally distinguished in any way from phonological rules proper. In previous approaches, the special status of redundant information had been expressed by first omitting redundant specifications from phonological representations, and then inserting these specifications *in the course of the derivation*, interleaving the blank-filling redundancy rules with the phonological rules proper. As a corollary, phonological rules had access to the redundancy free, UNDERSPECIFIED representations.

Stanley (1967) argued (1) that in order to avoid unwanted consequences in the interactions between the redundancy rules themselves, the rules must be severely constrained, so as to be (a) unordered and (b) monotonic (that is, prevented from changing feature values). In addition, he argued (2) that in order to avoid unwanted consequences in the interactions between the redundancy rules and the phonological rules proper, all redundancy rules should operate in a block, prior to the application of any phonological rule. As a consequence, phonological rules had access only to FULLY SPECIFIED representations. By the time representations enter the phonology proper, then, the distinction between redundant and non-redundant specifications is lost.

Stanley (1967:424) concludes that this particular constellation of properties that characterize redundancy rules (also known as Morpheme Structure rules, or MS rules) is most appropriately expressed in the following way:

the most natural way of formulating such statements is not in terms of rules at all, and we will show that the MS rules can be replaced by a new device, MS CONDITIONS; we use rules (like

the P rules) only to map one level onto another, and 'conditions' to state redundancies at a given level.

Stanley notes that such a formulation avoids all formal problems, and also allows us to state generalizations which are not easily stated in terms of rules.

Although Stanley rejects the notion of (blank-filling) redundancy rules, his proposed model still makes use of lexical representations ('dictionary matrixes') which omit redundant information — that is, his model still employs blanks. At first glance, it might appear that Stanley is thus faced with the problem of formulating a 'blank-filling condition', a 'procedural constraint'. Stanley, however, clearly distinguishes the procedure that executes the constraints from the constraints themselves.

There are two distinct parts to the system: a declarative part — the statement of the conditions — and a procedural part — a process which 'fills in the blanks'. The first part consists of:

an UNORDERED set of MS conditions which defines, in a manner specified in detail [below], a set  $M(U)$  of FULLY SPECIFIED matrices.

[...]

The second part of the proposed system is a PROCESS OF SELECTION... The incompletely specified dictionary matrix  $D_m$  SELECTS, in a manner described [below], a member of  $M(U)$ . (Stanley 1967:425)

Stanley could hardly be more insistent that there are two distinct components in this model, and insists that it is the *process of selection* that 'provides a method for filling in redundant information in dictionary matrices'. He remarks:

The statement of constraints and prediction of redundancies, though intimately related, are given as separate processes.

He is equally at pains to point out that the morpheme structure conditions *themselves* do not predict redundant feature values. They do not predict them, let alone insert them. The role of conditions in the grammar should not be taken to specify, fill, instantiate, or modify the lexical representations in any way. As we will see below, the formal role of such conditions is to ACCEPT fully specified matrices as well-formed.

We insist on this distinction, partly because Stanley himself is so insistent on it, partly because of its relevance to the argument we develop below, and



partly because it has been repeatedly collapsed by writers who claim to describe Stanley's position. Brown (1969), for example, writes:

Entries in the lexicon are in the form of non-redundant matrices. Redundancy rules ('morpheme structure rules' (Halle, 1959), 'morpheme structure conditions' (Stanley, 1967), 'lexical redundancy rules' (Chomsky & Halle, 1968)), assign redundant feature values to the output of the lexicon.

Statements such as this confuse the role of conditions and the role of redundancy rules, and thus the essential difference in function — and the essence of a constraint-based approach to well-formedness — is obliterated. As Stanley (1967:426) reminds us:

the FUNCTIONAL difference between an if-then condition and an MS rule should be kept clearly in mind. The former is a *statement* which defines a subset of the set U of fully specified matrices, namely the subset consisting of those matrices in P that it accepts. The latter is an *instruction* for filling in blanks in matrices with '+' and/or '-'. [my italics MB]

We now return to the role of MS CONDITIONS in the phonology, as conceived by Stanley, and the precise sense in which these conditions define the set of well-formed morphemes. His exposition runs as follows. First consider the closure U of all possible feature matrices for a language, where the language has n distinctive features, and in which the longest morpheme has  $\lambda$  segments (thus U is finite). An MS condition is a statement of a property shared by some, but not all, matrices, e.g. 'has two columns' 'begins with a [+consonantal] segment', 'has no two consecutive [-vocalic] segments'. Each condition, then, divides the set into two parts, one part consisting of all the matrices which have the property, and its complement. Stanley (1967:425) writes:

We will say that an MS condition C ACCEPTS a matrix M in U if M has the property stated in C, and that C REJECTS M otherwise

This way of approaching the notion of grammaticality is similar to the approach adopted in the mathematical definition of a relation, which often begins with the notion of a cartesian product, and then defines the relation in terms of a designated subset of that set. As Stanley (1967:428) notes, the set of MS conditions may thus be thought of as filtering out, from the set U of all matrices, those matrices which do not form possible morphemes of the language.



This then is the form and nature of MS conditions, as developed by Stanley. We note that it represents a constraint-based, non-derivational approach to well-formedness, explicitly distinguishing the statement of well-formedness from any procedure that might make reference to that statement, such as the procedure of determining the value of a blank in some environment. As we will see subsequently, the position we adopt with respect to redundancy obviates the need for such a procedure, and results in an exclusively constraint-based approach.

## 1.2. Syntactic well-formedness

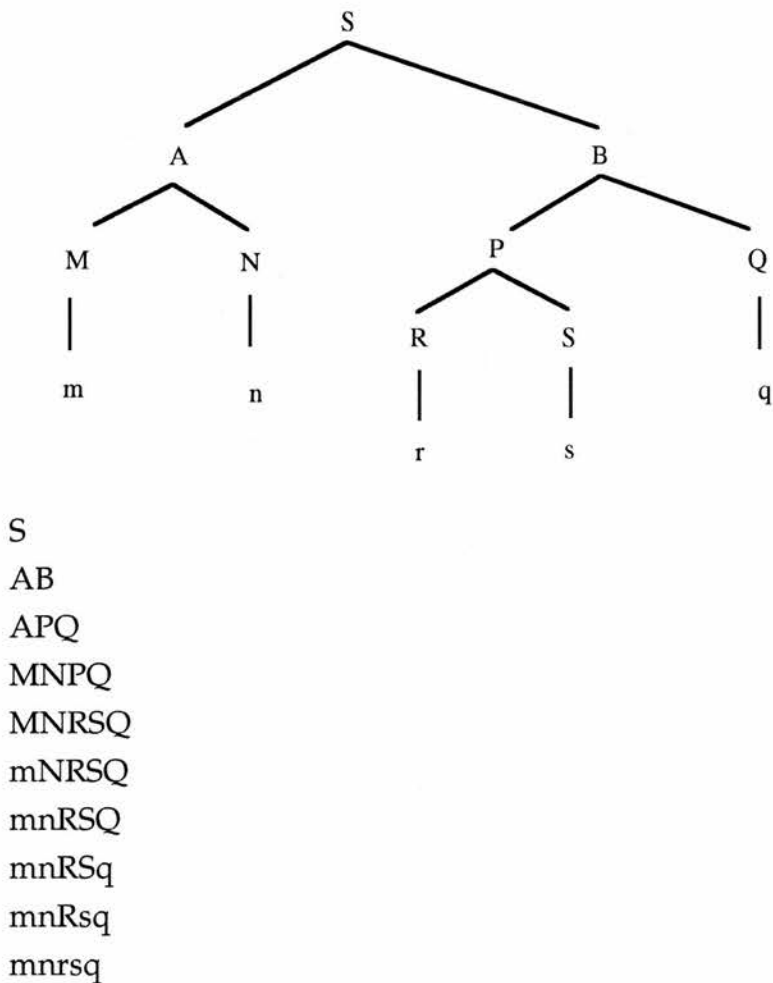
Through the advocacy of McCawley, Stanley's approach to the characterization of well-formedness was to have repercussions outside the domain of phonology, and to become extremely influential in the development of non-transformational and unification-based *syntactic* formalisms. The importance of this cannot be stressed too highly. The most radical difference between these formalisms on the one hand, and the transformational paradigm of the time on the other, concerns the basic organization of a generative grammar — its overall architecture, the very notion of a derivation. A crucial step in developing this alternative approach was introduced in McCawley (1968), and he attributes the idea to Stanley.

In this paper McCawley suggests an alternative formulation of the *base component* of a transformational grammar — including the phrase-structure component, which specifies the structures which transformations operate on. This alternative formulation is motivated by specific technical problems with the standard model. These specific problems need not concern us here, but it is important to understand what the nature of the standard model was, and the nature of McCawley's revision.

The particular formulation of the base-component which McCawley addresses had been most recently set forth in Chomsky (1965), and crucially involved the notion of REWRITING RULES. The term is borrowed from the mathematical theory of rewriting systems. A REWRITING SYSTEM is a finite set of rules  $\phi \rightarrow \varphi$ , where  $\phi$  and  $\varphi$  are strings of symbols. A DERIVATION in a rewriting system is a sequence  $x_1, x_2, \dots, x_n$  of strings such that each subsequent line is obtained from the preceding one by substituting according to the rewriting rules. The derivation is terminated when the last line consists solely of (pre-designated) TERMINAL symbols, and the last line is said to be

GENERATED by the rewriting system. McCawley notes that although mathematical rewriting systems contain only unordered sets of rules, Chomsky's earlier work had imposed an *extrinsic ordering* on the rules employed. Although this idea had not been exploited in mathematical work, the formalism did allow such a modification if it was felt, as Chomsky had, that specifically linguistic facts warranted it.

Further, if one makes certain restrictions on the form of the strings that are rewritten in the rules, one can define an algorithm which takes the derivation — the sequence of strings — and outputs the mathematical object known as a TREE. Again, the details of this algorithm need not concern us, but it is important to note that, in such a system, a phrase structure tree like the one in below is constructed, not directly from the phrase-structure rules, but from the intermediate stage of the *derivation* , also illustrated below:



As mentioned above, McCawley points to certain inadequacies of the rewrite rule approach, and proposes an alternative conception of the base component which involves neither ordering nor rewriting rules. He notes that certain difficulties can be obviated if the base component operates directly in terms of phrase-structure trees, rather than through the intermediate stage of a rewriting system. The motivation here is simply to make the apparatus less clumsy and more efficient, and to avoid technical problems which are simply artifacts of the formalism. McCawley considers two alternatives.

The first alternative (which he attributes to Stockwell et al. 1965) has this property. In this conception, the base component consists of TREE-FORMATION RULES (p.247):

a rule is interpreted not as an instruction to, say, replace a symbol *A* by a sequence of symbols *BC* but rather as an instruction to put two nodes labelled *B* and *C* under a hitherto terminal node labeled *A*. In place of a rewriting rule derivation (sequence of strings of symbols), this proposal substitutes a tree derivation: a sequence of trees

This approach, then, removes the problem of the indirect relationship between derivation and tree, but retains the notion of a derivation. It involves a list of instructions to sequentially modify parts of the tree, just as the rewrite system sequentially modifies parts of a string. That such a sequential derivation is intrinsic to this alternative is apparent from the fact that, just as in the rewrite rule approach, this proposal also allows the instructions to be extrinsically ordered.

It is the second alternative that is of most interest here, for in addition to meeting the specific inadequacies raised by McCawley, this proposal differs radically in its approach to well-formedness:

In the other proposal, to my knowledge first suggested by Richard Stanley (personal communication July 1965), the notion of 'derivation' is dispensed with entirely: the base component is a set of *node admissibility conditions*, for example, the condition that a node is admissible if it is labelled *A* and directly dominates two nodes, the first labeled *B* and the second labeled *C*.

It will be apparent that this proposal does for syntactic rules just what the proposal of morpheme structure conditions does for morpheme structure rules. And since the notion of a derivation is dispensed with, so the possibility of extrinsic ordering is removed (McCawley 1968:248):

the admissibility of a tree is defined in terms of the admissibility of all of its nodes, i.e. in the form of a condition which has the form of a logical conjunction

The notion of node admissibility conditions was adopted by Generalized Phrase Structure Grammar. As Gazdar (1982:137) writes:

There are many ways of interpreting the formalism of a phrase structure grammar but only two of these ways need concern us here. One way, adopted in *The Logical Structure of Linguistic Theory* (Chomsky 1975), interprets a phrase structure rule as a rewriting rule, a rule which maps strings into strings. Thus the rule:

$$S \rightarrow NP VP$$

is a function which maps strings of the form  $X-S-Y$  into strings of the form  $X-NP-VP-Y$ . The derivation of some terminal string is the set of all the strings that arise in the mapping from the initial symbol to that terminal string. Given certain restrictions [...] a tree may then be defined on the basis of a derivation.

The second way of interpreting PS rules, due originally to Richard Stanley (see McCawley 1968:39), is to treat them as node admissibility conditions. A node labelled  $S$  in a tree is admitted by the rule [above] if and only if that node immediately and exhaustively dominates two nodes, the left one labelled  $NP$  and the right one labelled  $VP$ . A tree is analyzed by the grammar if and only if every non-terminal node is admitted by a rule of the grammar. Under this interpretation, then, phrase structure rules are *well-formedness conditions on trees*. There is no notion of a derivation and it makes no sense to order the rules.

This notion of 'admissibility' was to become pervasive in GPSG: phrase structure conditions were simply the first of a whole body of constraints whose satisfaction was required in order to successfully *admit* a local tree (Gazdar et al. 1985).

In addition, the sets of constraints admitting local phrase structure trees in GPSG embody a strictly similar notion of well-formedness as that embodied in the 'pool of constraints' approach of such related formalisms as Lexical Functional Grammar (Bresnan 1982) and Head-driven Phrase Structure Grammar (Pollard and Sag 1987). Such formalisms share a number of assumptions about 'what a grammar formalism should do' (Shieber 1986), and these shared assumptions stem in large part from the constraint-based approach to well-formedness that they all adopt. This shared assumption was expressed in a shared formalism, a formalism specifically designed to be

responsive to the constraint-based approach — the formalism of unification grammar.

### **1.3. Unification-based syntax**

The unification-based approach represents the confluence of several streams of research, in computational linguistics, theorem proving, knowledge representation, and the theory of data-types (Shieber 1986). A family of syntactic formalisms has arisen which exemplify the unification-based approach: the Lexical Functional Grammar (LFG) of Bresnan and Kaplan; the Functional Unification Grammar (FUG) of Kay; the Generalized Phrase Structure Grammar (GPSG) of Gazdar, Pullum, Sag and Klein; the Categorical Unification Grammar (CUG) of Uszkoreit; the Unification Categorical Grammar (UCG) of Zeevat, Klein and Calder; the Head-Driven Phrase Structure Grammar (HPSG) of Pollard and Sag. What all these formalisms have in common is the notion that linguistic objects be characterized in terms of partial information structures (Pollard and Sag 1987:7). Typically, a number of such structures characterize a single linguistic object, each contributing a partial description of the object, and constraining its ultimate form. It is the operation of combining or unifying these partial descriptions into a single composite description that gives the framework its name. Crucially, too, these structures can be viewed as partially ordered according to their relative degree of informativeness. One structure is said to *subsume* a second if it contains a subset of the information found in the second. Intuitively, the less information a description contains, the less specific and the more general it is. The general case subsumes the specific.

### **1.4. Unification-based phonology**

Work in this paradigm has typically taken a coarse-grained view of phonology: this is natural, since the focus of the work was syntax and semantics. But recently, a number of researchers have been investigating the applicability of the unification-based formalism, and the constraint-based notion of well-formedness it embodies, to phonology (Bird 1991). In view of the history of the approach outlined above, this is particularly appropriate.

The consequences of taking such an approach to phonological well-formedness in general can hardly be exaggerated, since generative phonology is permeated by procedurality. On a strict interpretation of the constraint-



based approach, there is no extrinsic order, no operation precedes any other operation, there *are* no operations; nothing turns into anything else, nothing is modified, nothing is built; there are no structure building rules; nothing moves, nothing spreads; nothing is preserved, nothing is repaired; feature values don't change, *nothing* changes; nothing is delinked, or devoiced, nothing becomes aspirated; there are no processes; nothing is '-ized', not palatalized, or nasalized; nothing is inserted, nothing is deleted.

At first glance, it may appear impossible that such an apparently impoverished theory could possibly account for the attested diversity of phonological regularities. It is notable, however, that time and time again in the phonological literature procedural formalizations of phonological rules are accompanied by English glosses which do *not* involve procedurality. Consider, for example, the following English descriptions of Yoruba sound patterns provided by Archangeli and Pulleyblank (1989):

- a low vowel (always [-ATR]) is permissible both to the left and the right of a high vowel (p. 176)
- it is impossible to have a [+ATR] mid vowel to the left of a low vowel (p. 176)
- with respect to tautomorphemic sequences of mid vowels, the values of [ATR] must agree (p. 177)
- mid vowels may be either [+ATR] or [-ATR] when following a low vowel but may only be [-ATR] when preceding a low vowel (p. 178)

All of these statements are (informal) constraint-based accounts of the relevant phenomena. The constraint-based approach to well-formedness actually results in extremely natural formulations of phonological regularity.

It may be objected that, even if some such set of constraints could be made observationally adequate, they must nevertheless be hopelessly complicated, lacking the simplicity and elegance of a procedural account. Compare, for example, the instructions in Halle and Clements's (1983) textbook example on Klamath, using data and analysis from Barker (1963):

In his grammar, Barker makes the following assumptions:  
'Phonological rules are unordered. All rules apply simultaneously to underlying representations to derive surface representations'. Show how Barker's sets of rules can be simplified by abandoning these assumptions and assuming that phonological rules apply in order, each applying to the output of the preceding rule in the list of ordered rules

Objections of this are reminiscent of objections to phrase structure grammars, objections which were rebutted by researchers in GPSG. Their response was, essentially, that this is a technical problem, and as such is amenable to a technical solution. As Gazdar (1981:159) puts it — with respect to a syntactic phenomenon for which transformational rules were thought to be essential — ‘Phrase structure grammars can handle unbounded dependencies in an elegant and general way, provided that we exploit the resources offered by a complex symbol system and by the possibility of making statements about the set of rules that the grammar may employ’. Technical advances involving feature structures and abbreviatory devices have continued within the non-transformational, unification-based paradigm, and some of these will be introduced in subsequent chapters. In addition, recent developments in phonology itself — elaborating representational aspects of phonology at the expense of rule systems — have gone a long way towards making a constraint-based approach feasible. Thus a textbook in declarative phonology would read: ‘Show how Barker’s sets of rules can be simplified *without* abandoning his assumptions, by employing a richer notion of phonological representation, and exploiting the technical devices of constraint-based grammars’.

### 1.5. Morpheme structure rules revisited

The research paradigm of unification-based or ‘declarative’ phonology thus attempts to extend the domain of the constraint-based approach from morpheme structure conditions — as advocated in Stanley (1967) — to phonology proper; and to replace the procedural apparatus of phonological rules with a declarative alternative. However, current trends in phonology have actually taken a step backwards along this road, replacing morpheme structure conditions with morpheme structure *rules*. This constitutes a retreat from the constraint-based view in the very area of phonology in which it had been most comprehensively and rigorously established. And as we will see, there is a sense in which this is certainly a correct move.

As mentioned above, one of the reasons that Stanley rejected the use of rules (as opposed to conditions) in the statement of morpheme structure regularities was that giving phonological rules access to the kind of representations that result — representations containing blanks — has unwanted consequences. Specifically, it entails that rules will be able to make

what Stanley takes to be unwarranted generalizations. The 'third value' encoded by a blank allows the formulation of new and unmotivated phonological classes, stepping outside the possibilities of a binary system.

We may present the argument as follows: the use of blanks allows us to refer to the set of sounds which have '0' for some feature F as opposed to the set of sounds which have '+' for that feature (in exactly the same way that the theory of binary features allows us to refer to the set of sounds which have '-' for some feature F as opposed to the set of sounds which have '+' for that feature). Now for Stanley, a sound will have '0' for the specification of a feature just in case that specification is PREDICTABLE or REDUNDANT.

Therefore, the use of blanks allows us to refer to the set of sounds for which *F* is *predictable* as opposed to the set of sounds for which *F* is unpredictable. This is a class which is outside the set of natural classes which binary classification is intended to capture; and thus the use of blanks, and the accompanying rules which fill them, was rejected by Stanley.<sup>1</sup> Note, however, that Stanley's argument is against the *unwitting* reference to this class: using blanks allows reference to this class; reference to this class is undesirable; therefore, we must reject the use of blanks.

Recent work in underspecification theory, however, has shown that the set of sounds for which *F* is unpredictable is a natural phonological class: and that reference to such classes permits the simplification of phonological rules, and allows us to constrain rule-systems in desirable ways. Accordingly, work in this theory has *reintroduced* the use of blanks, precisely to refer to this class of

---

<sup>1</sup>Recall that Stanley does not claim that a phonology with blanks requires rules to fill them: on the contrary, he also shows that a process can be defined with respect to conditions in such a way as to provide a mapping between representations with blanks and fully specified representations. What Stanley shows is that giving *phonological rules* access to representations with blanks requires that the blanks be filled by rules. For what this means is that there is a stage of representation when the phonological rule applies, and at which a morpheme structure rule has failed (as yet) to apply. And one cannot translate this notion into one based on conditions, since the EXTRINSIC ORDERING of rules which it assumes is one of the possibilities (feature changing is the other) explicitly excluded by the conditional reformulation. Thus we can interpret Stanley as saying, not that blanks cannot be filled, but that the filling cannot be extrinsically ordered: it is this that results in three-way distinctions.



sound. And together with blanks go the concomitant morpheme structure rules which fill in the missing specifications in the course of the derivation.<sup>2</sup>

However, as we will see, quite independently of the issues which concern underspecification theory, there are a number of conclusive arguments *against* the use of blanks in phonological representations — against, that is, the decision to encode redundancy by omission. Moreover, the decision to use the apparatus of blanks and insertion rules leads to internal problems and incorrect predictions in underspecification theory itself. We will examine these arguments in detail in subsequent chapters: some have appeared at various points in the phonological literature; some we advance for the first time.

Thus we arrive at the following result: we need to be able to refer to the types of classes that underspecification theory makes available; but we cannot use underspecification to do it. However, simply because it is possible to refer to ‘the set of sounds for which F is unpredictable’ using the mechanism of blanks and morpheme structure rules, this does not mean that it is necessary to use this machinery. All that is required is an alternative, principled way of encoding the relevant distinctions.

## 1.6. Specification theory

This thesis proposes just such an alternative specification theory — an approach which we dub STRUCTURED SPECIFICATION. Our approach has consequences not simply for the issues addressed by underspecification theory, but for a whole range of phonological phenomena. An attendant advantage of the approach is that it is thoroughly constraint-based, preserving the notion that redundancy statements are best stated in terms of conditions involving logical implication. The main body of this thesis is concerned with establishing and motivating the main principles of this theory

---

<sup>2</sup>This issue has been complicated by the fact that many underspecification accounts (e.g. Kiparsky 1982, Pulleyblank 1983, Archangeli 1984, 1988) have taken the view that Stanley’s arguments provide compelling evidence against the use of ternary oppositions; thus they have attempted to argue that although they employ blanks in their representations, independent properties of their frameworks ensure that the use of blanks does not lead to ternary power. We argue that such a position is based on a misunderstanding of Stanley’s position, and fails to recognize that the extra classificatory capacity made available by a system with blanks is precisely the capacity that underspecification theories exploit.

of specification, and considering its consequences for other areas of phonological theory, and for the related fields of speech perception and recognition.

Our alternative approach will permit us to refer to the types of class identified by work in underspecification theory, but also preserves the notion that morpheme structure statements be expressed as *conditions* on representations, as opposed to *rules* that successively modify representations. Thus, empirically and theoretically, our approach supplies constraint-based phonology with an appropriate theory of specification.

The thesis is organized as follows. Chapters 2 and 3 exemplify in detail aspects of the declarative approach to phonology which provides our general framework. In Chapter 2 we illustrate the constraint-based approach to phonology by providing an account of phenomena traditionally expressed in terms of rule interactions, phenomena involving Grassmann's Law and Bartholomae's law in Sanskrit. The constraint-based approach is not only able to account for the interactions in a principled way, but avoids intractable ordering-paradoxes which befall a rule-based account. In Chapter 3, we illustrate the unification-based approach to phonology, and introduce aspects of unification formalism which will be crucial in later chapters. A recurrent theme of the thesis is the relevance of Firthian school Prosodic Analysis to the constraint-based approach to phonology. Indeed, we base these chapters in large part on specific analyses W. S. Allen, and in Chapter 3 show that the unification-based approach can trace its intellectual inheritance back to Firth through the work of Halliday in the framework of Systemic Grammar.

The following chapters form the core of the thesis, in which we propose an alternative treatment of phonological redundancy and markedness. In Chapter 4 we critically review the treatment of redundancy in generative phonology. In Chapter 5 we introduce the main principles of our 'structured specification' approach. In Chapter 6 we consider some immediate consequences of adopting such an approach, and in Chapter 7, we consider the implications of structured specification for the theory of feature geometry. In Chapter 8, we extend the approach to phonotactics in general, consider the model of the lexicon that results, and consider the implications for perception and recognition. In Chapter 9 we contrast our approach with that of underspecification theory. And in Chapter 10, we demonstrate that the

empirical predictions that follow from such a non-derivational approach to specification are borne out in a number of languages.

## 2. Constraint-based phonology

---

Two diachronic processes of Sanskrit — Grassmann’s law and Bartholomae’s law — present an ordering paradox when formulated as extrinsically ordered rules in a synchronic phonology. We argue that the paradox can be avoided by formulating the interaction in terms of simultaneous constraints on representations — an approach first suggested by Allen (1951). Like Allen’s, our analysis is constraint-based and non-derivational, stating the relevant generalizations in terms of templates and well-formedness conditions rather than rules. Essentially there is no ordering paradox because nothing is ordered.

### 2.1. Introduction

It has long been recognized that the interaction between Grassmann’s law and Bartholomae’s law causes problems for the synchronic phonology of Sanskrit. Kiparsky (1965) and Zwicky (1965) both noted that the standard formulation of these laws in a generative framework employing linearly ordered rules results in an ordering paradox. Both attempted to reformulate the rules in such a way that they might be accommodated by a linear ordering. Anderson (1969, 1970) criticized these approaches, arguing *inter alia* that the modifications proposed by Zwicky and Kiparsky require reference to unnatural segment classes. Rather than attempting to refine the structural descriptions of the rules involved, Anderson proposes to refine their mode of application, arguing for a theory of ‘disjunctive linear’ ordering.

None of these accounts locate the source of the problem in the decision to impose an ordering on phonological rules in the first place. But if the inclusion of a particular theoretical device results in paradox, then it surely suggests that the theory of grammar would be better without it. The challenge, of course, is to provide an explanatory account of the observed phenomena without ordering. Insofar as this is achieved, we provide a constructive critique of previous approaches.

The account adopted here eschews extrinsic rule ordering entirely, thus sidestepping the ordering paradox. The account draws on the leading ideas of

Allen’s (1951) treatment of Sanskrit aspiration — couched in terms of Firthian Prosodic Analysis — and takes its formalism from various current proposals in non-linear phonology, whose affinity with prosodic analysis has been repeatedly acknowledged. Where the formalism is modified, it will be seen to more closely approach a Firthian perspective — an incidental advantage of working the machinery of non-linear phonology in conjunction with an original example of Prosodic Analysis. Our account is formalized in terms of ASSOCIATION CONSTRAINTS and PROSODIC LICENSING. The analysis also makes use of the hierarchical organization of features advocated by the theory of feature geometry — relying crucially the presence of a laryngeal node dominating voice and aspiration features — and the notion that assimilation creates linked autosegmental structure. We argue that once these mechanisms are adequately formulated the behaviour of aspiration in Sanskrit to a large degree falls out automatically, with few language particular stipulations.

The chapter is organized as follows. In section 2 we discuss the data and previous approaches, setting out the nature of the paradox. In section 3 we outline the basic architecture and descriptive devices of Prosodic Analysis, seeking formal devices within current approaches to non-linear phonology which will appropriately express these insights within a generative framework. In section 4 we use this model to formalize and amplify Allen’s (1951) account of Grassmann’s law, continuing in section 5 with an account of the interactions with Bartholomae’s law. We conclude with some observations regarding the use of conventions and constraints in generative phonology.

## 2.2. Previous approaches

### 2.2.1. Grassmann’s Law

Grassmann’s law describes a diachronic process of aspirate dissimilation. If aspirates originally occurred in successive syllables of a word, and the second aspirate was part of the root, then the first occurrence of aspiration was lost. This dissimilation is evident in the reduplicated forms of aspirate initial verbs:

(1)

<b>*pha-phala</b>	→	<b>pa-phala</b>	‘burst’
<b>*ḍhu-ḍhauka</b>	→	<b>ḍu-ḍhauka</b>	‘approach’
<b>*cha-khaada</b>	→	<b>ca-khaada</b>	‘chew’

Dissimilation does not take place, however, if the second aspirate is not part of the root:

(2)

<b>bi-bhr̥-tha</b>	'bear' 2 PL. ACT. PRES.
<b>da-dhaa-the</b>	'put' 2 DU. MID. PRES.

The standard generative account is to treat this law as part of the synchronic phonology of the language, and to formulate a dissimilation rule which deletes the first aspirate. The following is taken from Anderson (1970:388):

(3) *Grassmann's Law* (GL):

$$[+cons] \rightarrow [-asp] \quad / \quad \text{---} \quad [+seg]_0 \quad \begin{bmatrix} +cons \\ +asp \\ +Root \end{bmatrix}$$

In addition to Grassmann's law, Sanskrit has an independent rule of deaspiration which applies before an obstruent or # boundary, and gives rise to the following alternations:

(4)

ACC. SG	NOM. SG.	INSTR. PL.	
<b>path-am</b>	<b>pat</b>	<b>pad-bhis</b>	'road'
<b>vṛdh-am</b>	<b>vṛt</b>	<b>vṛd-bhis</b>	'increasing'

This rule, which we will call Final Deaspiration, is formulated by Anderson (1970:388) as follows:

(5) *Final Deaspiration* (FD):

$$[+cons] \rightarrow [-asp] \quad / \quad \text{---} \quad \left\{ \begin{array}{c} [+obst] \\ \# \end{array} \right\}$$

Grassmann's law can also give rise to striking (synchronic) alternations. There is a class of roots which generally appear without initial aspiration, but do so just in case the final of the root is in a position to undergo deaspiration itself:

(6)

PRESENT	FUTURE	
<b>bodh-ati</b>	<b>bhot-syati</b>	'understand'
<b>duh-ati</b>	<b>dhok-ṣyati</b>	'milk'



Macdonell (1927:§55) describes this phenomenon as follows:

If *gh*, *dh*, *bh*, or *h* are at the end of a (radical) syllable beginning with *g*, *d*, *b*, and lose their aspiration as final or otherwise, the initial consonants are aspirated by way of compensation.

Such forms, which exhibit an apparently ‘mobile’ aspirate, are accounted for in a synchronic phonology equipped with rule-ordering by reconstructing the historical diaspirate form as the abstract underlying form, and extrinsically ordering FD prior to GL, in a bleeding relationship. Accordingly, certain roots will have an abstract underlying representation with two aspirates, although never appearing as such on the surface. This is exemplified by the following diaspirate root /**bhudh**/ ‘awakening’:

(7)

	ACC. SG	NOM. SG.	INSTR. PL.
U.R	<b>bhudh-am</b>	<b>bhudh</b>	<b>bhudh-bhis</b>
FD	—	<b>bhut</b>	<b>bhud-bhis</b>
GL	<b>budh-am</b>	—	—
	[ <b>budham</b> ]	[ <b>bhut</b> ]	[ <b>bhudbhis</b> ]

Under this approach, the apparent surface mobility of aspiration in such roots is simply a consequence of one or another of the underlying aspirate consonants being deaspirated.

As noted, this analysis relies crucially on the imposition of an extrinsic bleeding order:

1. Final Deaspiration
2. Grassmann’s law

The general methodology of such an approach within generative phonology was first suggested by Chomsky (1975:27, 29) in his treatment of Hebrew:<sup>1</sup>

It seemed only natural to construct a synchronic grammar with ordering of rules such as spirantization and reduction to explain the distribution of existing forms, and thus perhaps achieve the kind of explanatory force that I recalled from historical grammar. It became obvious that... I was being driven to more abstract underlying forms which were, in many cases,

---

<sup>1</sup>This quotation has been reorganized in order to preserve the sense.

historically more primitive as well. Furthermore, many of the rules suggested familiar historical processes.

The generative analysis of Grassmann's law thus constitutes a paradigm case of the explanatory adequacy of rule ordering. Consideration of Bartholomae's law, however, reveals that in certain instances the correct surface forms can only be obtained if the application of GL *precedes* Final Deaspiration, and to this we now turn.

### 2.2.2. Bartholomae's Law

Bartholomae's law modifies stop clusters, converting the sequence:

(8)

$$\begin{array}{cc} C1 & C2 \\ \left[ \begin{array}{c} +\text{voi} \\ +\text{asp} \end{array} \right] & \left[ \begin{array}{c} -\text{voi} \end{array} \right] \end{array}$$

into the sequence:

(9)

$$\begin{array}{cc} C1 & C2 \\ \left[ \begin{array}{c} +\text{voi} \\ -\text{asp} \end{array} \right] & \left[ \begin{array}{c} +\text{voi} \\ +\text{asp} \end{array} \right] \end{array}$$

Thus [dht] becomes [ddh], [bht] becomes [bdh], and so on, as in the following forms (Anderson 1970:388):

(10)

<b>ru(n)dh + tas</b>	→	<b>runddhas</b>
<b>ru(n)dh + thas</b>	→	<b>runddhas</b>
<b>labh + ta</b>	→	<b>labdha</b>

As Anderson points out, Bartholomae's law may be factored into two rules. The first is an assimilation rule which copies the voicing and aspiration features of C<sub>1</sub> to C<sub>2</sub>; the second is simply the rule of Final Deaspiration formulated above. Thus we have:

(11)

U.R.	<b>labh-ta</b>
ASS	<b>labh-dha</b>
FD	<b>lab-dha</b>



The ordering paradox surfaces when we consider diaspirate roots which undergo Bartholomae’s law, such as:

- (12)
- bhudh-ta**

‘awakening’ PA. PPL.

Applying the derivation as above, with Grassmann’s law bled by FD, we expect the following:

- (13)
- U.R.

**bhudh-ta**

ASS

**bhudh-dha**

FD

**bhud-dha**

GL

—
- \*[bhuddha]**

The problem is that Grassmann’s law is not in fact bled in this case: the required output is **buddha**, with initial deaspiration, which can only be achieved if GL is ordered *prior to* FD:

- (14)
- U.R.

**bhudh-ta**

ASS

**bhudh-dha**

GL

**budh-dha**

FD

**bud-dha**
- [buddha]<sup>2</sup>**

### 2.3. Prosodic Analysis

#### 2.3.1. Non-derivational phonology

Allen’s (1951) approach to Grassmann’s law is couched within the framework of Prosodic Analysis (PA), a non-linear framework, many of whose ideas have been incorporated into the theory of autosegmental phonology (Goldsmith 1979:203, Anderson 1985:192). However, as Anderson (1985:193) points out, a ‘major difference’ exists between the two theories:

Prosodic analysis... is an attempt to encode the effects of rules exhaustively within a theory of static representations. Autosegmental and metrical formalisms, on the other hand, are simply theories of the representations that appear within a

---

<sup>2</sup>The ordering of Assimilation relative to GL is not significant.

theory which contains significant representations that appear within a theory which contains significant rules as well... One consequence of this difference is that the extent of possible interrelation among 'prosodic' processes is much richer in this theory than in prosodic analysis, which has no analog of even simple cases of rule ordering.

This 'static', non-derivational organization of phonology is emphatically stressed by Sprigg (1965:1962) as being a central and distinguishing aspect of PA:

the basic-form concept, the general principle of treating any one form as basic, or as a norm, the remaining variants having only the status of derived forms, is repugnant to prosodic analysis. Firth's phonological theory is in this respect in marked contrast with theories that have adopted a derivational approach; and the contrast is greater now... for *derivation* and *process* now have the added support of generative grammars.

Clearly a theory with no analogue of rule ordering has no analogue of ordering paradoxes either, a positive advantage — provided the more limited theory is able to provide a satisfying account of phenomena which are generally explained in terms of extrinsic ordering.

It is striking that in the same year that Sprigg was reiterating Firthian dissent from the derivational approach in phonology, Richard Stanley was suggesting that the notion of derivation might be 'dispensed with entirely' in the field of syntax, as discussed in the previous chapter. It is worthwhile re-emphasizing this syntactic parallel, since the manner in which the idea was executed in generative syntax will provide us with concrete mechanisms with which to formalize the insights of PA in a generative framework.

Recall that Stanley's idea was first publicly advocated by McCawley (1968), who discusses in detail the relation between syntactic representations and the rules which determine their grammaticality. In *Aspects* and work previous to it, the well-formedness of an underlying representation or 'base-marker' derives from the fact that such a representation is emergent from some particular derivation in a rewriting system: a representation is grammatical just in case it emerges from some such derivation. The representation abstracts away from some trivial aspects of the derivation, but although the resulting phrase-marker is the object of primary linguistic interest in practice, it is nevertheless (formally) subservient to the derivation which gives rise to it. McCawley points to certain problems inherent in this manner of expressing

well-formedness, and traces their source precisely to the presence of the formal derivation mediating between rule and representation. He therefore argues that rules should state well-formedness over representations directly: that 'rules operate directly in terms of trees'. McCawley (1968:247) continues:

In the other proposal, to my knowledge first suggested by Richard Stanley (personal communication July 1965), the notion of 'derivation' is dispensed with entirely: the base component is a set of *node admissibility conditions*...

The most radical consequence of this proposal is that 'admissibility conditions' cannot be ordered with respect to each other. If one is *building* a tree — just as when one is building a house — doing things in different orders can give different results. If on the other hand one is merely checking that a tree is *well-built* or well-formed, then it doesn't matter in what order the checks are carried out: the structure remains the same. McCawley continues:

Node admissibility conditions are by nature unordered: the admissibility of a tree is defined in terms of the admissibility of all its nodes, i.e. in the form of a condition which has the form of a logical conjunction.

This conception of well-formedness informs all approaches to syntax which adopt some version of X-bar theory. Furthermore, syntactic rule-ordering was subsequently eliminated from the transformational component as well, whether by excision of the transformational component entirely, as in Generalized Phrase Structure Grammar (Gazdar et al. 1985), or its gradual contraction to the single rule Move- $\alpha$ , as in Government and Binding Theory (Chomsky 1981) — so that in current theory syntax is order-free. But this conception is still resisted within mainstream phonology. Indeed, Bromberger and Halle (1989) argue that it is precisely on these grounds that 'phonology is different'. As we have seen, however, the concept of an unordered, non-derivational phonology was explicitly advocated by proponents of Prosodic Analysis.

In attempting to incorporate the insights of PA within a generative framework, we are thus led to prefer the use of admissibility conditions over rules. These conditions may be positive (templates) or negative (filters), and are to be interpreted as statements of logical implication. It is no accident that Stanley (1967) is responsible for the introduction of just such devices into generative phonology, arguing for the inclusion of *Morpheme Structure Conditions* over *Morpheme Structure Rules*, and we will employ his notation

below. However, we go beyond Stanley in extending the range of phenomena such conditions properly describe, and in extending the relevance of admissibility conditions to the whole of the phonology — not merely the lexical specification of morphemes in isolation.<sup>3</sup>

### 2.3.2. *Systems and Prosodies*

The primary descriptive devices of PA have been succinctly described by Robins (1957): '*Prosodic Analysis* is, in fact, an abbreviated designation of an analysis that makes use of two types of element, Prosodies and Phonematic Units'. These two formal devices articulate respectively the syntagmatic and paradigmatic relations that obtain within phonological structures, as discussed by Firth (1957:5):

In dealing with language... two main sets of relations are set up, *firstly* the interior relations connected with the text itself. These subdivide into (a) the syntagmatic relations between elements of structure... (b) The paradigmatic relations of terms or units which commute within systems set up to give values to the elements of structure. For example, a five-term vowel system giving possible values for V in the first syllable of a CV<sup>5</sup>-CV<sup>7</sup>-CCV<sup>2</sup> structure.

Thus prosodies are established whenever phonetic features exhibit syntagmatic dependencies, and are stated over domains greater than a single segment: these prosodic domains may be strictly phonological (coda, rhyme, syllable, phonological word, etc.), or, 'since grammatically defined elements may also be characterized by prosodic features... we may have in addition word and morpheme prosodies' Robins (1957:4). We now express each of these notions using devices from current phonological theory.

- *Prosodies*

A leading idea of Prosodic Analysis is that, lexically, features may be 'unplaced' with respect to the string which realizes them, the surface position and extension of such features in phonetic substance being expressed by realization statements. In other words, while the *presence* of some phonetic feature in a morpheme may be unpredictable and form part of the lexical

---

<sup>3</sup>In this respect our approach has obvious parallels with proposals calling for the inclusion of 'output conditions' or 'surface structure constraints' in generative phonology, a position advocated by McCawley (1970) among others: see especially Shibatani (1973).



representation, the surface *position* of the phonetic feature is rule-governed, and does not form part of the representation. Such unplaced features are an example of what PA refers to as a PROSODY. The realization statements then determine the prosody's LINEAR PROJECTION, in Allen's felicitous phrase Allen (1951:941).

Such an insight was not formulable in the linear framework found in SPE and the type of phonemic analysis contemporary with Allen's work. In a linear framework presence is indistinguishable from position, and a property has to be placed in order to be represented. Allen (1951:940f.) explicitly criticizes this view in his treatment of the 'prosody of retroflexion' in Sanskrit:

Sanskrit spelling... is, in American terminology, largely 'phonemic' — that is to say, if a modern 'phonemicist' set out to produce an orthography for Sanskrit, the resulting spellings would show a close parallelism with those attested in Devanagari: as such, it is phonetically imprecise... and phonologically over-precise... It is thus in the tradition of the Sanskrit orthography that it should mark not only the presence of the prosody but also, to the best of its capabilities, its extension in the linear, phonematic dimension

Allen, however, 'adds a new dimension to the framework' (p. 945), and replaces the orthographic (and thus phonemic) representation **iṣṭa-** with the following prosodic representation:

(15)

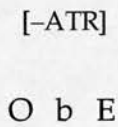
R

is - ta -

Current generative phonology is better equipped to incorporate this insight. Non-linear phonology crucially distinguishes between the CONTENT and STRUCTURE of phonological representations, the latter being represented formally by association lines. In such a framework we may simply say that the content of some representation is distinctive, but that its structure is not. The linear realization of the 'floating' feature, being predictable, will be left unspecified in lexical representation. General rules then determine how the feature links up with the CV-skeleton. Archangeli and Pulleyblank (1989:180) take this approach to the morpheme level prosody of [ATR] in Yoruba: 'To encode the property that nonredundant specifications of [ATR] are a feature

of morphemes, we propose that underlying [-ATR] specifications are unlinked... Underlyingly linked features, on the other hand, are required only when features are a property of a particular segment.’ Archangeli and Pulleyblank provide the following underlying representation of the morpheme **ɔbɛ**:

(16)



Which becomes, after general conventions and rules have supplied the structural relations:

(17)



Non-linear phonology thus provides the basis of an elegant formalization of the notion of a morpheme prosody, or morpheme-level feature.

- *Systems*

The second major device of PA is the ‘phonematic system’. The leading idea here is that the range of possibilities at a certain place in structure may be restricted to some subset of the possibilities of the language considered as a whole. Thus in Firth’s hypothetical example of a ‘CV<sup>5</sup>-CV<sup>7</sup>-CCV<sup>2</sup> structure’ cited above, a language with ‘a seven-vowel system’ as we should say, may in fact allow only a five-vowel subset to appear in the first syllable of polysyllabic words. Thus the notion ‘seven vowel system’ is in fact an abstraction from a phonological reality in which vowels in one structural position exhibit a five-vowel system, those in another a seven-vowel system, those in another a two-vowel system, and so on — in much the same way as the universal set of distinctive features is an abstraction over languages which employ determinate subsets of such features. The Firthian insight is that, just as the statements which express systematic gaps in the phoneme system of a particular language (the redundancy rules) may condition the application of

phonological rules in that language, so statements which express the systematic gaps in the phoneme system of a *particular structural position* in a language may condition the application of phonological rules in that position. The key point is that the phonology includes statements which constrain the set of oppositions available in particular structural positions. The structural position is termed a PHONEMATIC UNIT, and the range of oppositions operable there is known as a SYSTEM. These statements in turn interact with and constrain ('interfere with' in Allen's words) the prosodic realization statements in a modular fashion. That is, syntagmatic generalizations hold modulo paradigmatic constraints operating at particular positions, defined for the most part in terms of syllable structure.

The means of formalizing this insight within a generative framework are in fact at hand. Itô (1986) discusses constraints of just this type operating in a range of languages, and proposes a means of encoding them which she terms (appropriately enough in this context) PROSODIC LICENSING. Moreover, Itô formalizes this notion in terms of '*wellformedness conditions* on syllable representations rather than as part of the structural description of syllable-building rules' (p. 17, my emphasis).

An example of the kind of phenomena such conditions might be used to describe is to be found in Japanese. In many languages the range of oppositions operating in coda position is a subset of the oppositions which appear in onsets. As Goldsmith (1989:147) observes:

Firthian prosodic theory provides good terminological resources for describing this; Firthians refer to the (paradigmatic) range of possible segments in a given position in a syllable structure as a SYSTEM; this allows them to make the observation simply that the coda system is typically a subset of the onset system.

In Japanese, for example, obstruents are not permissible in coda position; nasals, on the other hand, are free to appear syllable finally. Thus we find:

(18)

\*kap.toot  
\*sek.pa  
\*kap.sek  
\*te.gak

but:

(19)

<b>sen.see</b>	'teacher'
<b>kam.pai</b>	'cheers'

Itô (1986:26) proposes to capture these facts by the following 'coda condition', which expresses a constraint on melody-CV association:

(20) *Itô's Japanese Coda Condition:*



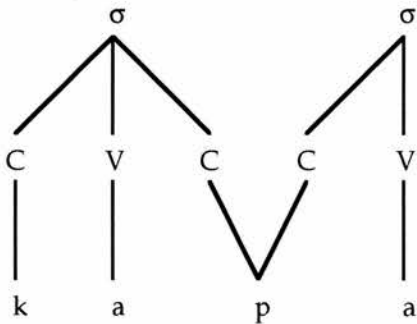
There is a systematic exception to the generalization noted above, however, involving geminate clusters: an obstruent *can* appear in coda position just in case the following syllable begins with an identical obstruent, as in the following forms:

(21)

<b>sek.ken</b>	'soap'
<b>gak.koo</b>	'school'
<b>kap.pa</b>	'legendary being'
<b>tos.sa</b>	'impulsively'
<b>toot.te</b>	'passing'

Examination of these geminate cases appears to show that Itô's condition is too strong as it stands. In the following form we do find a  $[-nas]$  segment associated with coda position:

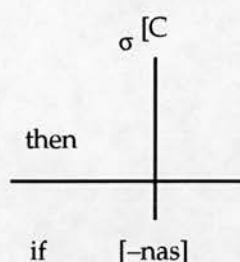
(22)





In order to accept such cases, Itô is forced to interpret her coda condition according to Hayes' (1986:331) Linking Constraint: 'Association lines in structural descriptions are interpreted as exhaustive'. We do not consider the merits of such a proposal here, but adopt an alternative approach. The generalization we wish to capture is that the linked structure of such heterosyllabic geminates in some way licenses the occurrence of the obstruent in the coda. In other words, association to coda position is licensed just in case the melody is also associated to onset position: it is thus association to *onset* position that is the necessary condition for well-formedness. We therefore propose the following positive well-formedness condition:

(23) *Japanese Obstruent Condition:*



That is, if an obstruent is present in the representation, it must be associated with an onset position. It follows that an obstruent will be licensed in a coda just in case it is associated with an onset: the desired result.<sup>4</sup>

This exceptional behaviour of geminates is not restricted to Japanese, but is a characteristic of many languages which exhibit coda restrictions, whether they be root geminates ('doubled' consonants such as **pp**, **tt** and **kk**) or partial geminates ('homorganic' consonants such as **mp**, **nt** and **ŋk**) (Prince 1984): this behaviour will be crucial to the analysis which follows. Note too that the Japanese Obstruent Condition (like Itô's coda filter) is (i) a well-formedness condition which (ii) expresses a relation between a set of possible segments

---

<sup>4</sup>See Bird (1990) for a similar proposal couched in terms of moraic theory. Itô's failure to consider this solution may be due to (i) her claim that 'a negative filter is formally equivalent to a positive wellformedness condition' (p. 32) (ii) her preference for stating conditions in terms of filters and (iii) the fact that no negative filter stated over onset position can achieve the desired result. However, the formal equivalence of positive and negative conditions holds only in the special case she discusses — where the consequent of the positive condition is Boolean. Itô's argument relies crucially on the fact that 'not [-nas]' is equivalent to '[+nas]'; but a moment's thought reveals that 'being in a coda' is not equivalent to 'not being in an onset', simply because — as is the case with geminates — a melody may be in *both*.

(picked out by the feature [-nas]) and a particular structural position (syllable onset). It is thus an entirely appropriate formalization of the Firthian notion of phonematic system. With these devices in place, we may now turn to the formalization of the prosodic approach to Grassmann's law.

#### 2.4. Allen's approach to Grassmann's Law

In roots which exhibit GL, Allen (1951) factors out aspiration as a prosody of the root morpheme, and provides a realization statement which applies modulo independent junctural constraints:

The prosodic treatment... notes simply that the aspiration is a prosody of the... radical syllable. As such, it requires to be marked [phonetically realized MB] once only... Generally marked by the special (aspirated) form of the second consonant, it becomes necessary to mark it elsewhere if, for reasons connected with interfering junction-prosodies, the second cannot carry it.

And in a later work, Allen (1973:10) provides a succinct statement of a strictly parallel phenomenon in a modern Indo-Aryan dialect, Harauti: 'a general descriptive rule can then be stated that in aspirated words the relevant phonetic feature occurs at, and only at, the first [in Sanskrit, last MB] possible location'. Allen (1951:944) goes on to provide the following representation of a Western Hindustani word, which exhibits 'a prosody of aspiration realized as breathiness of the vowel'. Allen's representation marks prosodic aspiration above what he describes as a CVVC structure:

(24)

H  
(b ə w t)

Such a representation may be incorporated directly into a non-linear framework - which as we have seen employs structures which are notationally and conceptually analogous - as follows:

(25)

[+asp]  
b ə w t

As noted above, according to current assumptions the realization of such morpheme level features is effected by linking the floating aspiration autosegment to the CV-skeleton. The kind of statement used to express these linkages is exemplified by the following universal ASSOCIATION CONVENTIONS taken from Archangeli and Pulleyblank (1989:181):

(26) Universal Association Conventions} (automatic)

Wherever possible, associate autosegments to anchors in a manner that is

- a. directional (left to right/right to left) and
- b. of a one-to-one nature.

The strictures on realization noted by Allen ('at and only at', 'last possible position') would seem to accord exactly with a right-to-left application of these conventions. Moreover, the use of such conventions seems particularly appropriate in view of the fact that, by definition, their application involves no extrinsic ordering: being automatic they are interpreted as 'applying continuously' throughout the derivation (Goldsmith 1976), that is, wherever and whenever their structural description is met. We will adopt these conventions for the moment for the sake of concreteness and familiarity, reserving the right to modify them later (section 6).<sup>5</sup> It is apparent that this aspect of the realization of the prosody of aspiration requires no language particular stipulation or rule.

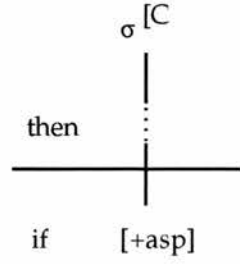
As Allen notes, however, there *are* language particular constraints on the realization of aspiration, which he attributes to 'interfering junction-prosodies', but does not analyze further. We now provide an explicit account of this 'interference' in terms of the conditioning influence of the different phonematic systems operating in onsets and codas.

Aspiration in Sanskrit is subject to the following condition:

---

<sup>5</sup>To anticipate the following arguments, it will be found that there is an unwanted procedurality inherent even in such 'conventions', and we argue below for an approach to association formulated directly in terms well-formedness conditions. However, it is important to note that the ordering paradox is avoided under either interpretation since neither conditions nor conventions can be ordered.

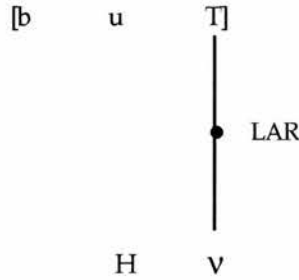
(27) *Sanskrit Aspiration Condition:*



That is, aspiration must be associated with onset position. Note that this is simply a declarative, non-derivational reformulation of Anderson's rule of Final Deaspiration (5 above). This condition constitutes an explicit statement of the 'possible locations' for the realization of the prosody of aspiration: formally, it interacts with and constrains the universal association conventions. The lacuna in the association line expresses the fact that [+asp] does not dock in the C slot directly; rather, we adopt the standard organization of feature geometry according to which aspiration is grouped (together with voice) under a laryngeal node (Clements 1985). This organization of aspiration and voice is crucial to the analysis which follows.

The lexical representation of the 'diaspirate' root is as follows, with an unassociated, morpheme-level aspiration prosody:<sup>6</sup>

(28)

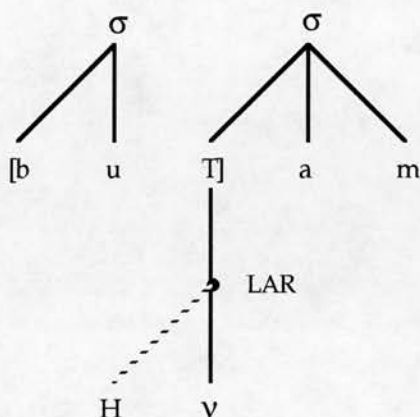


The unmarked or 'elsewhere' pattern of realization is found in the form budham. When concatenated with the vowel initial suffix **-am**, the final of the root is syllabified as an onset. H associates to (the laryngeal node of) this final

<sup>6</sup>In the following diagrams irrelevant structure is suppressed. 'H' denotes [+asp], 'v' denotes [+voi], H and v are on separate planes.

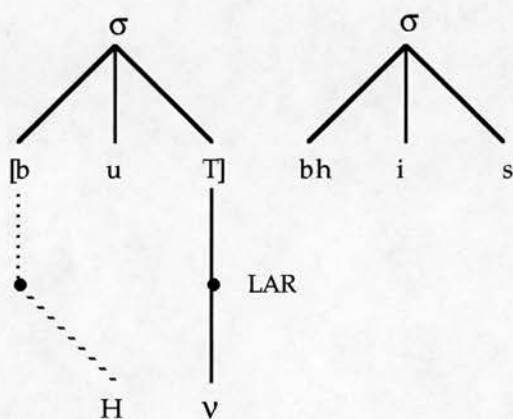
segment in accordance with the association conventions, and in conformity with the Aspiration Condition, yielding a [-asp]... [+asp] pattern in the root:

(29)



When the final of the root is not in onset position, however, the Aspiration Condition rules it out as a legitimate anchor: the autosegment accordingly is associated with the next eligible position, yielding the alternative [+asp]... [-asp] pattern. This is exemplified by the form **bhudbhis**, with consonant initial suffix **-bhis**:

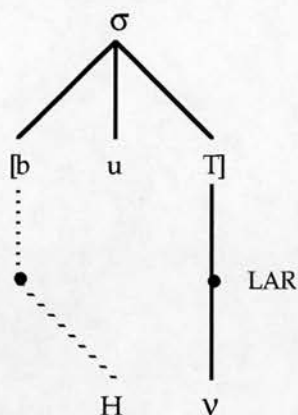
(30)



The same pattern emerges in the unsuffixed form **bhut**, where the final of the root is again in coda position (as above, we ignore final devoicing):



(31)



This then completes the analysis of Grassmann's law. It is not necessary to set up underlying forms for diaspirate roots which include an abstract sequence of aspirates — a sequence which is impossible on the surface, and which is obligatorily and destructively modified by rules operating in a fixed order. Under Allen's approach, which involves no feature-changing or extrinsic ordering, the apparent mobility of aspiration is simply a consequence of (i) the requirement that aspiration be marked at some one point in the root, which is (ii) conditioned by the independent phonematic constraint on association. The observed behaviour is thus modelled in terms of a set of interacting constraints, rather than an ordered sequence of operations.

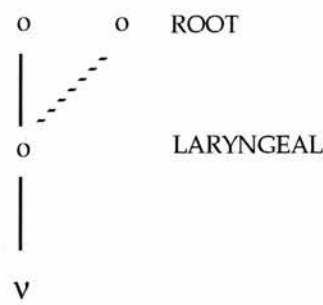
## 2.5. Interaction with Bartholomae's Law

In this section we argue that the exceptional behaviour of Grassmann's law in the context of Bartholomae's law follows from the same kind of interaction of constraints on representations. The observed pattern of aspiration in such forms will be found to be the only one capable of satisfying all the relevant constraints on well-formedness simultaneously. Since well-formedness conditions are by definition unordered, an account of interaction couched in terms of constraint satisfaction cannot be subject to ordering paradoxes.

As noted above (2.2) the problematic case for the derivational approach is **buddha**, where Bartholomae's Law has applied to the heteromorphemic cluster [dh-t]. From our standpoint, it is crucial to note that Bartholomae's law involves voicing of the succeeding stop. We formulate this as Laryngeal Assimilation:

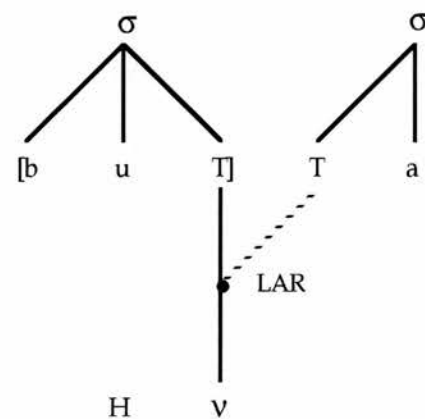


(32) *Laryngeal Assimilation (LA)*:



Just as in the case of the association of aspiration, LA may be interpreted as an association convention which applies automatically whenever its structural description is met, and thus cannot be extrinsically ordered.<sup>7</sup> In Firthian terms, LA expresses the linear projection of the laryngeal 'junction-prosody'. Crucially then, in words where LA applies, a partial geminate — a LARYNGEAL GEMINATE — will be created across syllable boundaries, creating a linked structure between the trigger (in coda position) and the following onset:

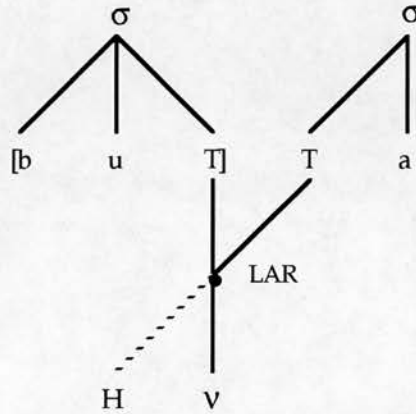
(33)



<sup>7</sup>See Archangeli and Pulleyblank (1989:187) for a discussion of the treatment of assimilation rules as conventions.

In just this case, then, aspiration (which, as noted above, docks in the laryngeal node) will be licensed in coda position of the root. Accordingly, association of the prosody to the right edge of the root will conform to the Aspiration Condition, yielding the [-asp]... [+asp] pattern as desired:

(34)



Two aspects of this analysis require comment. Note firstly that our formulation of Bartholomae's law as LA, unlike Anderson's formulation, does not mention aspiration explicitly. This is deliberate. We assume that the 'source' of the aspiration in Bartholomae's law is twofold. On the one hand, it may originate (as above) in the floating feature, which, to use the derivational metaphor, associates 'after' LA has created the necessary linked structure: this is the situation that obtains in the diaspirate roots we have been considering. On the other hand, aspiration may be an inherent feature of the final segment of the root itself, as appears in monoaspirate roots such as **path** 'road', which do not undergo Grassmann's law. In such cases the aspiration feature is associated with the final segment lexically, and — again using a derivational metaphor — is spread together with voice, subsumed under the laryngeal node. In either case, the same surface structure is obtained. In our non-derivational approach, of course, this is viewed as a 'conspiracy' which is explained by representations being required to satisfy simultaneous constraints — Grassmann's Law, the Aspiration Condition, and Laryngeal Assimilation — which hold of root final juncture in Sanskrit.

Secondly — again contra Anderson — we do not formulate Bartholomae’s law as the linear *transfer* of aspiration from one position to the next. In this respect our assessment of the orthographic transfer of aspiration is strictly parallel to Allen’s criticisms of Sanskrit’s orthographic representation of retroflexion. A Firthian would describe our analysis as involving ‘a junction-prosody of voicing and aspiration’, represented by the following formula:

(35)

$$\begin{array}{ccccccc} & & & & H & & \\ & & & & v & & \\ b & & u & & T & - & T & & a \end{array}$$

### 2.6. Conventions or Conditions?

The analysis above has avoided extrinsic ordering, expressing generalizations in terms of association conventions and well-formedness conditions. These two devices — conditions and conventions — have been yoked together ever since Goldsmith’s original Goldsmith (1976:27) definition of the well-formedness of autosegmental representations:

(36) *Goldsmith’s ‘Well-formedness Condition’*:

1. All vowels are associated with at least one tone: all tones are associated with at least one vowel.
2. Association lines do not cross.

It is important to note that, despite the name ‘Condition’, Goldsmith does not regard (36) as a declarative statement of admissibility, but rather remarks that ‘the Condition is interpreted so as to *change the representation* minimally so as to meet the condition maximally’ [p. 27, my emphasis]. According to Goldsmith the Well-formedness Condition creates (37b) from (37a) by ‘the addition of four association lines’:

(37)

a. archipelago

|

H L

b. archipelago

H L

In subsequent work, such as Clements and Sezer (1982:218-9), these structural changes are explicitly distinguished from the well-formedness conditions

they enforce. Under this formulation, the Association Conventions constitute an algorithm which manipulates representations so as to ensure satisfaction of the (independently stated) well-formedness constraints:

The present theory further assumes a set of universal Association Conventions which implement the Well-formedness Conditions... These conventions act as 'monitoring' devices in phonological derivations to preserve well-formed patterns of association between P-segments and P-bearing units.

Although the form of these association conventions has been subject to constant revision in the literature (see Pulleyblank 1986:9ff. for an overview), their function — an implementation algorithm mediating between statements of well-formedness and representations — has been accepted without comment. From a non-derivational point of view, however, it is immediately apparent that this model of well-formedness incorporates an unwanted layer of procedurality mediating between rules and representations. The alternative, in the spirit of Stanley's notion of admissibility conditions, is to interpret well-formedness conditions as admitting or rejecting structures directly. Under this interpretation, (37b) is admitted and (37a) is simply rejected. Representations which fail to meet the well-formedness conditions are not 'repaired', but filtered. It is worth quoting van Riemsdijk and Williams (1986:157) on the use of such a filtering strategy in syntax:

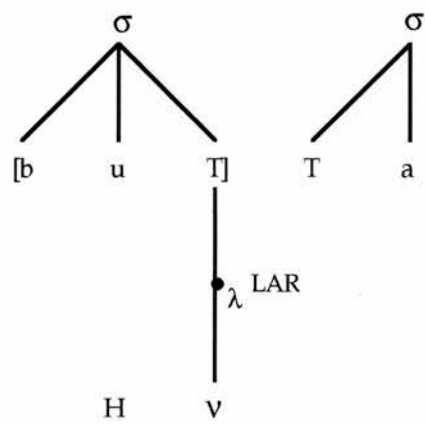
The introduction of filtering devices into the model of grammar has sometimes drawn misconceived criticism. In particular, some have felt that it is a priori wrong to first generate a wide range of structures only to filter out most of them later. But the appeal of this view is inspired by a procedural interpretation of the model. If the model were a production model... excessive reliance on filtering devices could (indeed, should) be frowned upon. However, since the model linguists are developing is a competence model... no such considerations apply. The only valid criteria for evaluating filters, like any other component of the grammar, are the degree to which they correctly characterize and explain grammatical phenomena and the degree to which they add to the overall elegance of the theory of grammar.

A careful examination of the way in which association interacts with the Aspiration Condition in Sanskrit reveals that the filtering interpretation of association statements is in fact the preferred interpretation, and that it is desirable (and, given the non-derivational aspect of the theory, appropriate) to model Allen's prosodic realization statement in terms of a declarative

ASSOCIATION CONSTRAINT rather than as a procedural association convention.<sup>8</sup>

Consider the following representation:

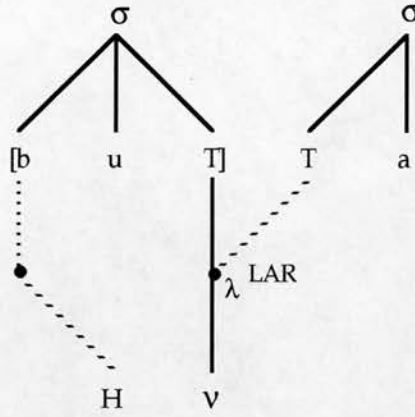
(38)



If we interpret association procedurally in the manner of Goldsmith (above), and as we have assumed thus far, then two association lines have to be introduced into this representation: one between H and the laryngeal node  $\lambda$ , the other between  $\lambda$  and the unassociated T. If these additions are unordered with respect to each other, then it is possible that H may attempt to dock in *before* the geminate structure has been created which would license it. In Archangeli and Pulleyblank’s 1989 account of Yoruba [ATR] harmony, a similar situation is found, with a floating [–ATR] feature attempting to dock on a vowel specified as [+hi] — a position where it is not licensed. In their account ‘the [–ATR] specification skips the high vowel... and links to the rightmost eligible bearer of a [–ATR] specification’ (p. 184). However, if in the above example H were to ‘skip’ to the next eligible bearer at this point in the derivation, to be followed subsequently by LARYNGEAL association, we would derive the incorrect form \*bhudda:

---

<sup>8</sup>Note, in this regard, Steriade’s (1988:286) observation on Allen’s approach to accentuation in Ancient Greek: ‘An attempt at formalizing Allen’s ideas reveals that the central element of the system is not a rule, tonal or accentual, but a filter ruling out tonal configurations in which the contonation is followed by more than one syllable.’



At this point we might propose that ‘skipping’ is a language particular parameter, and that association, while iterative in Yoruba, is non-iterative in Sanskrit, and simply fails if it can’t succeed. In the example above, this yields the right results. The initial attempt at association would fail, but a subsequent attempt — which would be made by virtue of the convention’s continuous mode of application — would ultimately succeed, once LA had created the requisite structure. The problem, of course, is that we *do* want association to skip to root onset position in certain forms in Sanskrit, in order to produce words with the [+asp]... [–asp] pattern of aspiration, such as **bhut** and **bhuddhis**.

The problem here lies entirely in the excessive procedurality of the mechanisms employed. If, on the contrary, we interpret association conditions directly as surface structure constraints, all difficulties are removed. We thus interpret association conditions to say that ‘the following pattern of association is well-formed’ rather than ‘create the following pattern of association’. This is the spirit in which Allen’s realization statement ‘the relevant feature occurs at, and only at, the last possible location’ is to be taken: not an instruction to position the feature at the last location, but to rule admissible those representations which conform to the pattern, and rule out those which don’t.

The formalization of such a constraint takes seriously the proposal that admissibility conditions have the form of ‘statements of logical implication’. In the next chapter, we develop such a formalization utilizing the tools developed in the field of logic-based approaches to syntax and semantics.



### 3. Unification-based phonology

---

Classical SPE phonology makes a rigid distinction between rules and representations, which are organized in an ordered series of input-output mappings. Contra SPE, the technique of Prosodic Analysis, practised by J. R. Firth and members of the London School, attempts to subsume the effect of rules *within* a theory of representations, organized in a modular fashion as an unordered set of constraints or 'prosodies'.

Functional Unification Grammar, developed by Martin Kay (1979), is a declarative, computationally tractable grammar formalism, which deliberately blurs the distinction between rules and representations: the same data-structure is used for both. In Unification Grammar a 'rule' is construed as a partially specified representation, a representation is just a fully specified rule. We argue that UG is thus an appropriate framework in which to explain and formalize the fundamental principles of Prosodic Analysis. As illustration, we reconstruct Allen's (1957) treatment of aspiration in Harauti within this framework.

Finally, we argue that the suitability of Unification Grammar to Prosodic Analysis is no accident, and trace a continuous line of development from Firth to Kay via the mediation of Halliday's work in Systemic Grammar.

#### 3.1. Introduction

Generative phonologists have frequently been bemused by the London school of Prosodic Analysis (PA). Anderson's (1985:179) observation is typical:

Firth's writings on general linguistics (and on phonology in particular) are nearly Delphic in character. Even the papers one might most expect to present systematic expositions of his theoretical position, such as Firth (1957), are full of obscure and allusive references and completely unclear on essential points.

And Chomsky (1964:75) observes:

I have been unable to find formulations of [the London school's] position that are explicit enough to show what evidence might count either for or against them.

If the general formulations of the London school seem obscure, the particular analyses presented by its members are notable for their attention to phonetic detail and the range of languages they treat, and the theoretical terminology employed is meticulously systematic and consistent. The perceived lack of definition, then, may be due more to the instrument of observation than the object observed. In a theoretical framework like that of classical generative phonology, with its rigid distinction between rules and representations, it is impossible to form a clear picture of what a 'prosody' is. But such a distinction is simply inappropriate to the type of grammatical organization envisaged by prosodic theory. This chapter argues that there is an explicit, formal theory of grammatical objects — Unification Grammar (UG) (Kay 1983) — in which the distinction between rules and structures is deliberately blurred, and in which the profile of PA emerges distinctly. The major aim of the chapter is to provide a formally explicit definition of 'prosody' in unification terms.

The chapter is organized as follows. In section 1, I explicate Allen's (1957) treatment of Harauti aspiration, translating his analysis into a more familiar, theory-neutral notation. In section 2, I introduce the necessary apparatus from Unification Grammar that will be used, in section 3, to model Allen's analysis. Finally, in section 4, I suggest that the suitability of unification to PA is no accident, and sketch a continuous line of development from Firth to Kay via the mediation of Halliday's work in Scale and Category and Systemic Grammar.

## **3.2. Allen's analysis of Harauti**

### **3.2.1.**

Prosodic Analysis is concerned to describe regular syntagmatic dependencies between various classes of sounds in specified domains; it is a theory of the syntax of phonological structures. At one extreme, a prosody may amount to little more than the abstraction of a single feature from some cluster of segments. But prosodies are not confined to statements of mere phonetic continuity in a more or less extended domain. Prosodies, then, differ from long-components (Harris 1951) and autosegments (Goldsmith 1976) in two immediately apparent ways. First, they typically describe the distribution of a *set* of features, rather than simply involving a single feature and its associations. And secondly, prosodies are stated with reference to a

hierarchical organization of structural units. Both these attributes of PA are exemplified in Allen’s (1957) ‘Aspiration in the Harauti Nominal’.

Harauti is a Rajasthani language spoken in the Kota district of India. It displays the following consonant inventory:

bilabial	p	ph	b	bh	m	mh		
dental	t	th	d	dh	n	nh	ll	lh
alveolar							r	s
retroflex	ɽ	ɽh	ɖ	ɖh	ɳ		ɭ	ɭ
palatal	c	ch	j	jh				
velar	k	kh	g	gh				
glottal							h	

In common with most Indo-Aryan languages, Harauti has a series of stops with murmured release, in addition to the three-way contrast between voiced, voiceless aspirated and voiceless unaspirated stops. Phonetically, these series may be classified as follows (Ladefoged 1971:13):

		Voice Onset		
		voicing throughout	voicing immediately after release	voicing considerably after release
Glottal Stricture	voiceless	bh	p	ph
	voiced		b	

We will adopt the following phonological specification:

		aspiration	
		–	+
voice	–	p	ph
	+	b	bh

Although we have recourse to the familiar feature ‘aspiration’ here — translating Allen’s ‘breathiness’ — it must be carefully distinguished from the ‘aspiration’ of Allen’s title. The *prosody* of aspiration describes the syntagmatic dependencies between voicing, ‘occlusion’ (continuancy), aspiration in the narrow sense, and nasality, within the domain of the word. Allen (1957:72)

summarises the ‘rather complex syntagmatic relations’ which the prosodic statement is to account for as follows:

- i) A breathy transition is never followed or preceded by another breathy transition within the word.
- ii) A voiceless articulation is never followed by a breathy transition except immediately (i.e. when the voiceless articulation in question forms the prior term of the transition).
- iii) Within the above restrictions, breathy transitions from voiceless articulations are found either initially or non-initially; but by (ii) such non-initial transitions imply that no other voiceless articulation precedes.
- iv) Breathy transitions other than from a voiceless articulation are only found initially.

Allen distinguishes the ‘transitional’ feature breathiness from ‘coarticulatory’ features such as nasality and ‘occlusion’, observing that phonetically it is unclear whether the aspiration is post-consonantal or pre-vocalic. He further observes that it is in the spirit of PA not to arbitrarily associate the feature with one or other segment, but rather to ‘abstract the transition from the sequence as a whole’. We will model this below as a minimal prosody of phonetic continuity embedded within the larger prosody of aspiration.

We now provide our own summary of these constraints. All aspirates are bound by the following condition:

- i) only one  $\begin{matrix} C \\ [+asp] \end{matrix}$  per word

In addition, voiced aspirates are bound by the condition:

- ii)  $\begin{matrix} C \\ [+asp] \\ [-voi] \end{matrix}$  must be first C

and voiceless aspirates by the condition:

- iii)  $\begin{matrix} C \\ [+asp] \\ [-voi] \end{matrix}$  must be first  $\begin{matrix} C \\ [-voi] \end{matrix}$

Historically, these syntagmatic dependencies arise by a process of dissimilation like that of Grassmann’s law. Thus, Prakrit **bhikkha** ‘alms’ > H.

**bhik** (the second aspirate feature being suppressed in accordance with (i) above) and Prakrit **pokkhara** 'lotus pool' > H. **phokar** (the aspiration being thrown back to the voiceless initial in accordance with (iii)).

As mentioned above, PA describes dependencies between sounds in specified domains. A prosody is essentially a statement of licit co-occurrences of sounds in specific structural positions. PA argues that a level of syntagmatic structure is essential to such a statement. Part of this programme is the elevation of traditional but informal terms of description — terms like ‘initial’, ‘medial’ and ‘final’ — to an explicit, formal status. But there are other relevant domains as well, which may be nested inside one another in a hierarchical structure of considerable subtlety. We now reconstruct this level of structure.

### 3.2.2.

Allen writes (p. 72):

The exponents of the prosodic units are referable to certain places in structure; these reference points are provided by

- (a) single C elements;
- (b) 2-element sequences of CV, CC, C(I), or (F)V (where I = initial of following word and F = final of preceding word... )

Working 'bottom-up' then, these 'elements and sequences' constitute the first level of syntagmatic structure imposed on the string. In terms of (b) above, a simple string such as:

(1) CVCVCV

will be grouped into the 2-element sequences:

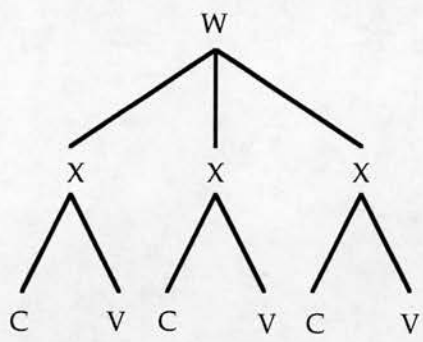
$$(2) \quad [C \ V][C \ V][C \ V]$$

In tree form<sup>1</sup>:

<sup>1</sup>The following abbreviations are used in the diagrams: W(ord), Fo(cal), Pr(efocal), Po(stfocal), C(onsonantal), V(owel).



(3)

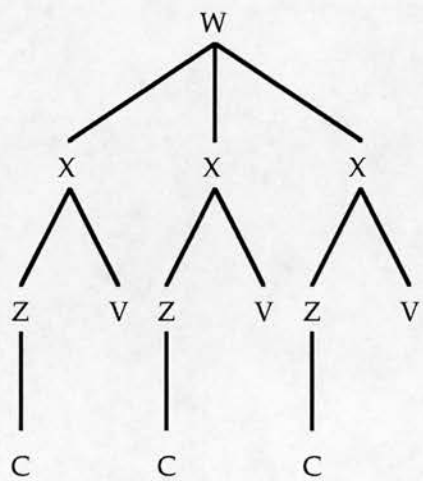


With respect to Allen’s point (a) it is essential to grasp that the ‘single C’ elements are not exclusive of the 2-element sequences, but included within them, thus:

(4)                                    [[C] V] [[C] V] [[C] V]

Again in tree form:

(5)



This will become clearer when we examine the next level of structure, where these sequences themselves are grouped into units.

A prosody is a statement of licit co-occurrences of sounds. Allen argues that four mutually exclusive sets of licit co-occurrences are possible with respect to the cluster of features comprehended under the prosody ‘aspiration’. Each of these possibilities is designated a ‘term’ in the prosody. Thus the prosody of aspiration is a four-way exclusive disjunction of terms, where each term can

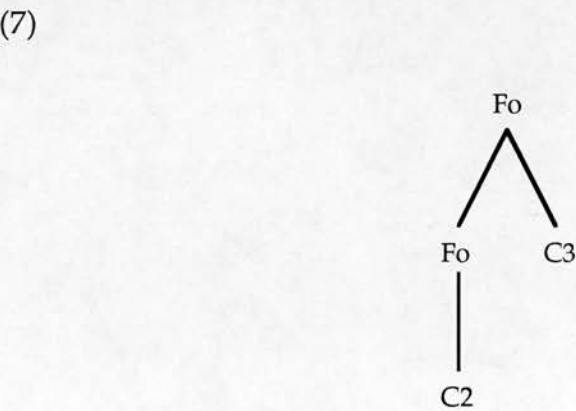
be thought of as a kind of template. To take a syntactic analogy, the prosody of subject-verb agreement in English is a statement of the licit co-occurrences of the features [number] and [person]: each such possibility is a term. The syntagmatic structure with reference to which these terms are stated is not uniform, however. Two terms of the prosody of aspiration are stated with respect to 'such key positions as initial, medial and final'. But for the statement of the other two terms these reference points are not sufficient. For these, Allen (1957:73) argues:

A particular 2-element sequence is indicated (by means of a subscript ligature) and is termed 'focal'; of the two elements the first is also termed focal. By orientation to this sequence and element other reference points — of type (b) or (a) — may be stated as 'prefocal' or 'postfocal'.

It is these terms we will concentrate on in this chapter. Note that *both* the sequence and its first element are termed focal. This first element is the 'single C' element referred to in (1) above. Allen illustrates this level of structure by example.

(6) C1 V1 C2 C3 V4

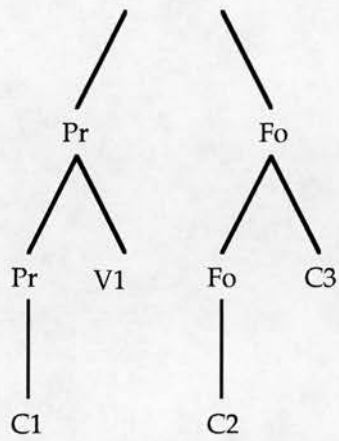
Here 'the sequence C2 C3 and the element C2 are focal' (p. 73). Allen clearly implies that both a sequence and an element within it may be considered focal *within the one structure*. We may represent the structure of this sequence as follows:



This type of 'headed' structure is familiar from X-bar theory: the focal constituent contains a focal element within it, just as a verb phrase contains a head verb. (Since the structures described by Allen are not recursive, there is evidently a notion of bar-level implicit here, which we suppress in our

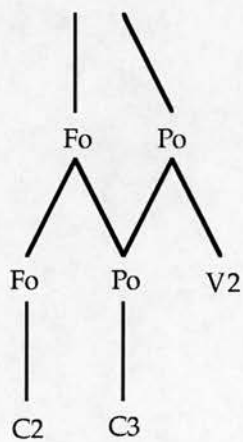
representation). Allen continues ‘the sequence C1 V1 and the element C1 are prefocal’ (p. 73) again stressing the headed structure. We thus have:

(8)



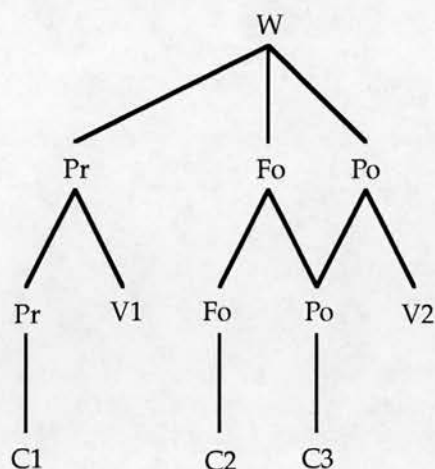
The postfocal material is not merely the vowel V2, but quite explicitly ‘the sequence C3 V2 and the element C3’ (p. 73). This gives rise to an ‘ambisequential’ daughter dominating C3, where one and the same element is simultaneously the non-head element of the focal constituent and the head of the postfocal constituent:

(9)



This type of structure is reminiscent of the ambisyllabicity of Kahn 1976. The final structure, then, will be:

(10)



Note too that this level of structure is not isomorphic with syllabic structure, treating as it does both CV sequences and CC clusters on an equal footing.

3.2.3.

If this is an example of the kind of structure utilized by PA, what is its grammar? What is the possible range of structures, and how are they induced? Allen summarises the attested structures in terms of the following templates (p. 74):

(11) CV(...)

(12) ...CV(...)

(13) ...CC...

(14) ...C(I)

It seems that Allen has been inconsistent with his abbreviatory conventions here. (12) evidently abbreviates the following two structures:

(15) ...CV

(16) ...CV...

But a systematic use of this standard convention allows us to collapse (12) and (11) as follows:

(17) (...)CV(...)

Furthermore, (17) bears a marked similarity to (13), the only difference being that the material preceding and following a consonant cluster is obligatory: a consonant cluster is not permitted to begin or end a word. This is clearly due to independent constraints on syllable structure, and not to be considered part of the statement of the distributional constraints on aspiration. Accordingly we propose to factor out this information and make it part of a separate statement. Note that such

move is quite in the spirit of PA itself. Allen observes (p. 70):

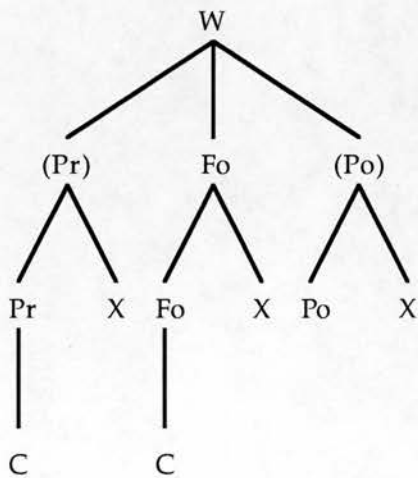
the structure to which one prosodic system is relevant may overlap or include that to which another is relevant — an incomplete prosodic analysis involves the allotment to phonematic units of data which in a complete analysis would be allotted to other prosodic units

In this case, we argue, Allen allots data to a *prosodic* unit (aspiration) which in a fuller analysis would be allocated to another prosodic unit (syllabification, say). We thus collapse (13) and (17) as (18) employing an partially specified segment X, and assuming that an independent (prosodic) statement will further specify X as C only in cases where the cluster is non-peripheral.

(18)                                    (...)CX(...)

This schema, then, abbreviates (11-13) above. (For the purposes of exposition we put aside the sandhi case (14)). Translating this into a form consistent with the notation we have adopted for structures in (10), we can say that words must conform to the following template:

(19)





This then is the hierarchical syntagmatic structure with reference to which the prosody of aspiration is stated.

3.2.4.

We are now in a position to examine the prosodic units themselves.

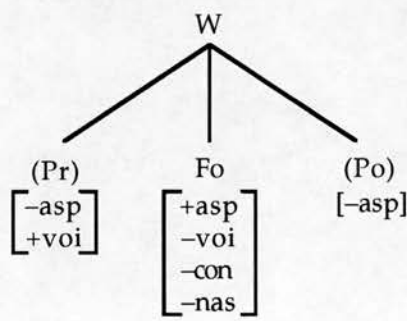
Here is Allen’s account (p. 77):

Statement of prosodic units...

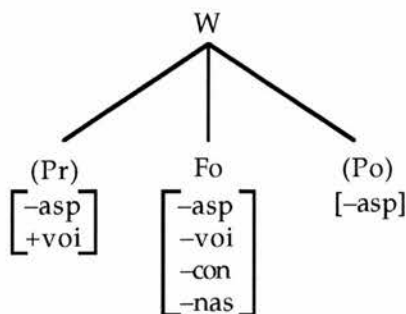
- h:
- Transitional:
    - Focal: [breathiness]
    - Non-Focal: [non-breathiness]
  - Coarticulatory:
    - Focal: [voicelessness occlusion non-nasality]
    - Pre-focal:[voice]
- h’:
- Transitional: [non-breathiness]
  - Coarticulatory:
    - Focal: [voicelessness occlusion non-nasality]
    - Pre-focal:[voice]

This is the statement of licit co-occurrences of features, made with reference to the syntagmatic positions identified earlier. We can express this economically by annotating the template (19) with the appropriate features:

(20) h:



(21) h':



Although it is not immediately apparent from Allen's formulation, it is clear that these two prosodic terms differ only in respect of the focal constituent's value for aspiration.

We can capture the distinction between transitional and coarticulatory feature behaviour by employing a distinction familiar from work in the GPSG framework Gazdar et al (1985). We note that coarticulatory features turn up systematically on what we have identified as the head segment. Call these 'head' features, subject to the following constraint:

(22) *Head Feature Convention (HFC)*:

The head features of the mother must be an extension of the head features of the head daughter.

On the other hand, transitional features are properties of 'the sequence as a whole'. Call these 'foot' features, subject to the following constraint:

(23) *Foot Feature Principle (FFP)*:

The foot features of the mother must be identical to the foot features of every daughter<sup>2</sup>.

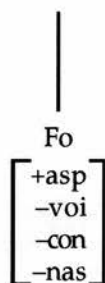
These conventions interact with the prosodic template and headed hierarchical structures outlined above. As an example, recall that the focal constituent of the template (20) is:

---

<sup>2</sup>This is a stronger version of the FFP than that found in GPSG, where it is sufficient that one daughter bear the foot features of the mother.

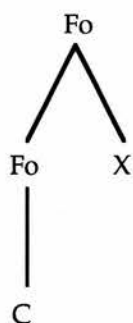


(24)



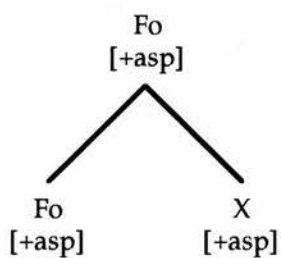
while the internal structure of the cluster was formulated as:

(25)



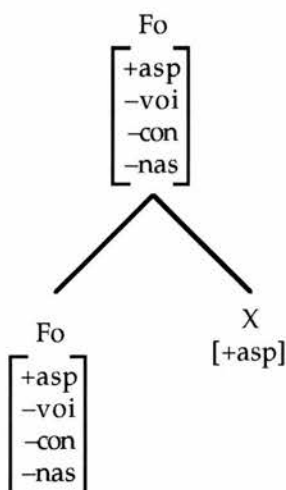
The FFP ensures that the transitional feature [aspiration] is distributed within the structure as follows:

(26)



And in addition, the HFC is responsible for the mother-daughter agreement of coarticulatory features on the head projection:

(27)



### 3.3. Rules, representations and unification

#### 3.3.1.

As indicated earlier, there has been some confusion in the generative literature as to whether prosodies should be regarded as rules or representations. On the one hand, we have a categorical statement like the following from Anderson(1985:189):

it is quite clear that the Firthian prosodic analysis is entirely a theory of representations... Regularities are incorporated into an analysis exclusively in the form of definitions of representational elements

But Anderson also observes that 'the representations in question are quite different from those countenanced by phonemic theories' (p. 189), and that:

the syntagmatic dependencies described in Firthian terms as prosodies are just the sort of regularity that generative descriptions would capture in terms of rules (p. 185)

Chomsky (1965) chapter 4, 'The nature of structural descriptions', is almost entirely devoted to phonological structure. In section 4.2 he outlines his theory of the phonological component as an input-output device, consisting of a set of rules mapping from systematic phonemic representations to the output systematic phonetic level. In the next section 'Taxonomic phonemics', where this celebrated term is coined, he compares and rejects the approach to sound pattern practised by contemporary structuralists. But note that for Chomsky, PA is *not* considered a taxonomic theory. In a footnote at the beginning of this section he explicitly dissociates it from such theories:

the prosodic analysis of the London school... in particular seems to have certain relations to the point of view sketched above in 4.2

For Anderson, prosodies are representations which nevertheless 'do the work' of rules. For Chomsky, PA — this theory of representations — nevertheless bears conceptual similarities to a rule-governed derivational theory.

The problem here is that the distinction inherent in the vocabulary being employed to talk about PA are not appropriate to PA itself. PA does not carve the world into mutually exclusive classes of rules and representations — the very distinction obscures the outline of the theory.

This is where Unification Grammar is useful.<sup>3</sup> UG gives us a formal vocabulary for describing linguistic objects which does not enshrine this distinction. In UG, rules and representations are not different kinds of object: a 'rule' can be construed as a partially specified representation, a 'representation' is just a fully specified rule.

3.3.2.

In UG, then, one and the same data structure subsumes both rules and representations. Such a structure is termed a **attribute-value matrix** (AVM). An AVM can be thought of as an elaboration of the familiar phonological feature-matrix; AVMs express information uniformly in terms of features and values. The elaborations are of two kinds, both of which concern the *kind of value* a feature may take. Firstly, rather than restricting each feature to an atomic value (typically, + or -), features may take entire matrices as values. Thus in (25) the feature CATEGORY takes the atomic term 'np' as value:

(28)

$$\left[ \begin{array}{l} \text{category} = \text{np} \\ \text{agreement} = \left[ \begin{array}{l} \text{number} = \text{singular} \\ \text{person} = \text{third} \end{array} \right] \end{array} \right]$$

---

<sup>3</sup>Unification Grammar (Kay 1979, 1983, 1985) has been known variously as 'Functional Grammar' and 'Functional Unification Grammar'. It was the first of the unification-based approaches to grammar, and uses unification as its only operation (see Shieber 1986).



The feature agreement, on the other hand, takes as value (26):

(29)

$$\begin{bmatrix} \text{number} = \text{singular} \\ \text{person} = \text{third} \end{bmatrix}$$

which is itself a feature-matrix.

Secondly, two distinct features may share a *single* value. In the following matrix, the feature [agreement] is not associated directly with a value, but we are told that its value is identical to some other feature's:

(30)

$$\begin{bmatrix} \text{category} = \text{np} \\ \text{agreement} = \langle \text{verb agreement} \rangle \end{bmatrix}$$

The value for this np's agreement is the same as the verb's value for agreement (whatever that may be) — these two features share one and the same value. Here the effect is similar to that obtained by using alpha notation in phonology — the matrices in (31), for example, are notationally equivalent:

(31)

$$(a) \begin{bmatrix} \alpha_{\text{con}} \\ \alpha_{\text{voi}} \end{bmatrix} \quad (b) [\text{con} = \langle \text{voi} \rangle]$$

The behaviour is not identical, however. The variable ' $\alpha$ ' ranges only over atomic values, and there is no unification equivalent of something like (32):<sup>4</sup>

(32)

$$\begin{bmatrix} -\alpha_{\text{con}} \\ \alpha_{\text{voi}} \end{bmatrix}$$

### 3.3.3.

Now consider the following AVM:

---

<sup>4</sup>See however, Coleman (1991) for a recent proposal to extend the formalism of unification to allow such categories.

(33)

$$\left[ \begin{array}{l} \text{pattern} = (\text{subj pred}) \\ \text{cat} = \text{s} \\ \text{subj} = \left[ \begin{array}{l} \text{cat} = \text{np} \\ \text{case} = \text{nom} \end{array} \right] \\ \text{pred} = \left[ \begin{array}{l} \text{cat} = \text{verb} \\ \text{subcat} = \text{intr} \\ \text{agr} = \langle \text{subj agr} \rangle \end{array} \right] \end{array} \right]$$

(The feature 'pattern' has a special interpretation: the order of its values corresponds to the concatenation order of the constituents). This AVM is a statement at about the same level of generality as the following rule:

(34)

$$S \rightarrow \left[ \begin{array}{l} \text{NP} \\ \text{nom} \\ \alpha \text{agr} \end{array} \right] \left[ \begin{array}{l} \text{V} \\ \text{intr} \\ \alpha \text{agr} \end{array} \right]$$

It describes an object whose syntactic category is 'sentence', with two constituents, subject and predicate, of category noun phrase and verb respectively, and whose agreement value is identical; it also specifies information about case and subcategorization. The AVM represents a partial specification of the properties of a certain class of sentences. Thus in UG a 'rule' can be seen as a partial representation, a partially specified AVM.

Unification Grammar gets its name from the operation by which objects become more fully specified. Two compatible AVMs can be combined to yield a single AVM which contains all the information of its inputs. This process is termed **unification**. Thus unifying (35) and (36) yields (37):

(35)

$$\left[ \text{subj} = \left[ \begin{array}{l} \text{cat} = \text{np} \\ \text{case} = \text{nom} \end{array} \right] \right]$$

(36)

$$\left[ \begin{array}{l} \text{cat} = \text{np} \\ \text{agr} = \left[ \begin{array}{l} \text{num} = \text{sg} \\ \text{per} = 3 \end{array} \right] \end{array} \right]$$

(37)

$$\left[ \begin{array}{l} \text{subj} = \left[ \begin{array}{l} \text{cat} = \text{np} \\ \text{case} = \text{nom} \\ \text{agr} = \left[ \begin{array}{l} \text{num} = \text{sg} \\ \text{per} = 3 \end{array} \right] \end{array} \right] \end{array} \right]$$

Unifying information about the NP in (36), information about the verb in (38) below, and the AVM (33) which corresponds to a sentential rule, we get the output (39):

(38)

$$\left[ \begin{array}{l} \text{cat} = \text{verb} \\ \text{voice} = \text{active} \\ \text{aspect} = \text{prog} \end{array} \right]$$

(39)

$$\left[ \begin{array}{l} \text{pattern} = (\text{subj} \text{ pred}) \\ \text{cat} = \text{s} \\ \text{subj} = \left[ \begin{array}{l} \text{cat} = \text{np} \\ \text{case} = \text{nom} \\ \text{agr} = \left[ \begin{array}{l} \text{num} = \text{sg} \\ \text{per} = 3 \end{array} \right] \end{array} \right] \\ \text{pred} = \left[ \begin{array}{l} \text{cat} = \text{verb} \\ \text{subcat} = \text{intr} \\ \text{agr} = \left[ \begin{array}{l} \text{num} = \text{sg} \\ \text{per} = 3 \end{array} \right] \\ \text{voice} = \text{active} \\ \text{aspect} = \text{prog} \end{array} \right] \end{array} \right]$$

If (33) is an AVM with the same level of generality as a rule, (39) is an AVM with the same level of specificity as a phrase-marker. Yet in UG both are precisely the same type of formal object: both rules and representations are uniformly conceived of as constraints on expressions of the object language (in this case, constraints on sentences). A 'derivation' consists of the incremental addition of information, resulting in a AVM which is fully 'saturated'. We can think of a representation as just the limiting case in which a set of constraints becomes maximally specific. In a similar way, when a

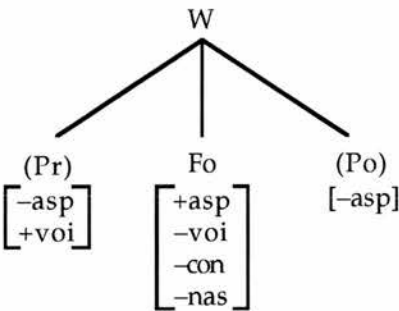
traditional phonological feature-matrix is maximally specified it is associated with a particular sound, rather than a class of sounds: it can then be interpreted as a representation of that sound.

### 3.4. Modelling prosodies by unification

#### 3.4.1.

UG, then, gives us just the kind of vocabulary we need to provide a formal account of a prosody. The AVM corresponding to the prosodic term (20) (repeated below as (40)) is (41):

(40)



(41)

$$\left[ \begin{array}{l}
 \text{pattern} = (Pr \quad Fo \quad Po) \\
 \text{cat} = W \\
 Pr = \left[ \begin{array}{c} \text{asp} = - \\ \text{voi} = + \end{array} \right] \\
 Fo = \left[ \begin{array}{c} \text{asp} = + \\ \text{voi} = - \\ \text{con} = - \\ \text{nas} = - \end{array} \right] \\
 Po = \left[ \text{asp} = - \right]
 \end{array} \right]$$

Recall that a prosody is a *disjunction* of terms, in this case differing solely in the focal value of aspiration. Disjunctions are easily expressed in UG: alternants are enclosed in braces ('{}'), which are interpreted to mean that just

one of the enclosed values must be unified with. A verb like **take**, for example, which may be either first or second person singular, or the plural of any person, would receive the following partial specification:

(42)

$$\left[ \begin{array}{l} \text{cat} = \text{v} \\ \text{agr} = \left\{ \left[ \begin{array}{l} \text{num} = \text{sg} \\ \text{per} = \{1, 2\} \end{array} \right] \right. \\ \left. \left[ \text{num} = \text{pl} \right] \right\} \end{array} \right]$$

We may thus collapse the two terms of the prosody in one AVM:

(43)

$$\left[ \begin{array}{l} \text{pattern} = (\text{Pr} \ \text{Fo} \ \text{Po}) \\ \text{cat} = \text{W} \\ \text{Pr} = \left[ \begin{array}{l} \text{asp} = - \\ \text{voi} = + \end{array} \right] \\ \text{Fo} = \left[ \begin{array}{l} \text{asp} = \{+, -\} \\ \text{voi} = - \\ \text{con} = - \\ \text{nas} = - \end{array} \right] \\ \text{Po} = [\text{asp} = -] \end{array} \right]$$

The AVM above *is* the ‘prosody of aspiration’.<sup>5</sup> Given our general discussion of the interpretation of AVMs, it is easy to see why it is difficult to characterize such an object as *either* a rule *or* a representation, and why a rigid adherence to such terms obscures its true nature.

This prosody may then be further specified. Unifying the AVMs in (44) with the prefocal, focal and postfocal constituents of the prosody (43) respectively, we obtain (45):

---

<sup>5</sup>Recall that this discussion is confined to just two terms of the complete four-term prosody.



(44)

$$\begin{array}{lll} \text{(a)} \begin{bmatrix} \text{place} = \text{lab} \\ \text{con} = - \\ \text{nas} = - \end{bmatrix} & \text{(b)} \begin{bmatrix} \text{place} = \text{vel} \end{bmatrix} & \text{(c)} \begin{bmatrix} \text{place} = \text{alv} \\ \text{voi} = + \\ \text{con} = + \\ \text{nas} = - \end{bmatrix} \end{array}$$

(45)

$$\begin{bmatrix} \text{pattern} = (\text{Pr} \text{ Fo} \text{ Po}) \\ \text{cat} = \text{W} \\ \text{Pr} = \begin{bmatrix} \text{place} = \text{lab} \\ \text{asp} = - \\ \text{voi} = + \\ \text{con} = - \\ \text{nas} = - \end{bmatrix} \\ \text{Fo} = \begin{bmatrix} \text{place} = \text{vel} \\ \text{asp} = + \\ \text{voi} = - \\ \text{con} = - \\ \text{nas} = - \end{bmatrix} \\ \text{Po} = \begin{bmatrix} \text{place} = \text{alv} \\ \text{asp} = - \\ \text{voi} = + \\ \text{con} = + \\ \text{nas} = - \end{bmatrix} \end{bmatrix}$$

If (43) is a formal model of a prosody, the AVMs in (44) model **phonematic units**. We see that the prosodic ‘rule’ (43) is a proper subpart of the ‘representation’ (45).

**3.5. Unification grammar and systemic grammar**

How is it that a grammatical formalism, designed by Kay in the late seventies to be a mathematically rigorous and computationally tractable approach to

syntax, should be so appropriate for modelling a type of *phonological* analysis developed thirty years earlier with neither of these aims in view? This is in fact no accident. As Kay (1979:152) acknowledges:

there is a strong family resemblance between grammatical descriptions in this formalism and the *systems* [of] Halliday (1961).

Halliday's Systemic Grammar came to the attention of computational linguists through Winograd's celebrated implementation in the early seventies (Winograd 1972). Kaspar (1985:1) shows how the devices of Systemic Grammar may be encoded in UG, as a means of making SG formally explicit:

if we look beyond differences of format, we find that UG and SG make many of the same assumptions about language and grammar. The similarities are due, in part, to the fact that when Martin Kay (1979) formulated UG he was responding to many of Michael Halliday's ideas

We refer the reader to this paper which shows in detail the intertranslatability of the key aspects these two formalisms.

But just as UG may be seen as a formalization of Halliday's approach to linguistic structure, so Halliday (1961) constitutes the first attempt to codify and formalize the Firthian approach: indeed, the school which developed from Halliday's seminal paper was dubbed 'neo-Firthian'. Every major theoretical term there defined is taken from earlier formulations by Firth. As Butler (1985:13) observes:

We must... admit that [Firth's] ideas were presented only programatically, often obscurely, and almost never with the degree of rigour which we have come to expect from modern linguistics. It was up to Halliday to take up the task of using Firth's ideas... to build a linguistic theory in which the categories and their relationships would be made explicit

Influence is transitive. There is a clear tradition from Firth to Kay via Halliday, characterized by increasing standards of formal rigour. It is quite natural to see a unification-based approach to Prosodic Analysis as the most explicit version of Firth's conception of linguistic structure developed thus far. Kay has supplied a formal model which, unlike that of mainstream generative phonology, is appropriate to this conception.

## 4. The treatment of redundancy in generative phonology

---

### 4.1. Introduction

Generative phonology has always included formal devices for encoding the notions of predictability, redundancy and distinctiveness, although at different times in its history different emphasis has been placed on the importance of such statements in the functioning of the phonology as a whole. Attention has recently been refocussed to this area by work in UNDERSPECIFICATION THEORY, which has established the conditioning role of such considerations in the operation of phonological rules proper, not just in the area of phonological inventories and phonotactic constraints. As the name implies, underspecification theory attempts to exploit the descriptive possibilities that follow from giving phonological rules access to representations in which certain features are omitted — the redundant, and in some approaches, the unmarked features.

Such underspecified representations have formed part of the apparatus of generative phonology from the very beginning. In fact, in view of the radical revisions to the theory of representations that have been made within the standard model — which now incorporates prosodic structure, autosegmental structure, and the structure of feature geometry — the notion that, at some stage or other, phonological representations are underspecified has remained uniquely unchallenged.

This thesis challenges the assumption. Central to the approach we will propose is the idea that redundant features, and unmarked features, are present at all stages of the derivation, a theory of ‘radical overspecification’. There is no ‘omission’ and ‘insertion’, redundant information is (potentially) available to all rules and constraints. Our approach makes redundant information available, but explicitly codes it *as* redundant, a coding which we derive from the structural relations which hold between distinctive and redundant specifications. In terms which have become familiar in current phonology, it is a representational, as opposed to a rule-based, approach to redundancy.

The chapter is structured as follows. We first critically review the treatment of redundancy in generative phonology, paying particular attention to the original model of Halle (1959). From this we derive a sense of the leading ideas behind the generative treatment of redundancy, and also re-evaluate the role of a largely forgotten means of executing those ideas, the BRANCHING-DIAGRAM. We also focus on Stanley's (1967) approach to redundancy; in particular, his final conjecture — a conjecture which was incorporated into neither his own model nor the model of SPE — regarding the role of blanks in phonology. In a sense, it is an analogue of the branching-diagram, informed by Stanley's final conjecture, that forms the basis of the formal model we develop in subsequent chapters.

#### 4.2. The problem of redundant information

A convenient introduction to the treatment of redundancy in generative phonology is provided by Anderson (1985:10ff.). In the introductory chapter of his history of twentieth-century phonological theory, Anderson seeks to provide a concrete example of the following thesis:

In order to really appreciate the logical content of our own [contemporary] views, then, it may be necessary, somewhat paradoxically, to approach this task through a prior appreciation of their historical antecedents... We must 'get inside' the position within which some problem originally arose in order to understand its motivations and logical underpinnings

Conveniently for our purposes, the particular issue Anderson chooses to 'get inside of' to illustrate his point is the following:

It is often taken as self-evident in phonological studies that underlying ('phonemic' or 'phonological') representations should contain only *distinctive* or nonredundant material. That is, in arriving at the phonological representation of a form, one of the steps involved is the elimination of all predictable properties, and the reduction of the form to the minimum of specification from which all of its other properties can be derived by general rule. For many, indeed, such a step establishes the fundamental difference between the 'phonological' and the 'phonetic' representation of a given form.

It is precisely this position that we argue against, and we shall do no better by way of introduction than summarizing Anderson's discussion of the 'intellectual inheritance' of this position in modern phonology.



Anderson's presentation of the issue has the following outline. After setting out the position in the terms quoted above, he immediately shows that the elimination of redundancy may have undesired consequences: specifically, a problem arises when forms have *reciprocally dependent* properties. He exemplifies this point using facts from Russian vowel-consonant interdependencies, and concludes:

In the worst case, we may be forced to make choices that cannot be defended on principled grounds just in order to meet the requirements of eliminating redundancy

Moreover, he observes that such a state of affairs is quite frequent in phonology; indeed, 'the simplest case of this type [...] occurs so frequently that it is not generally even noticed' (p.11). We will discuss his examples of reciprocal dependency below, and note that this issue has recently been raised again by Mohanan (1991), who makes precisely the same point.

Having demonstrated the problems, Anderson next defends the logical coherence of an opposing view, in which at least some predictable information is retained underlyingly:

the position that such representations should be redundancy-free is not self-evidently correct [...] it is perfectly possible to develop a view of phonological forms which is consistent with the fundamental function of these representations in a grammar, but in which at least some predictable detail is present

Anderson also makes the observation that some researchers in speech sciences work *solely* in terms of representations which are maximally specified.

In the face of the manifest problems of the standard view, and given the coherence of an alternative approach, 'How', asks Anderson, 'did the position arise in which it is all and only *un*-predictable properties that appear in phonological representations?' He identifies two potential sources of this position; one arising from an interpretation of Saussure, the other arising from information theory. On one interpretation of Saussure, the units of phonological structure are *identified with* sets of properties distinguishing them from other units. Anderson comments:

The doctrine that 'dans la langue, il n'y a que le différences' has often been interpreted as equating the phonological character of a sound with exactly those properties that distinguish it from



others — no less but no more. Thus, there would be no room in such a representation for properties that were not distinctive

However, argues Anderson, this is just one interpretation of Saussure's position; one which owes more to theory-internal issues of the structuralism espoused by Saussure's successors and interpreters than to Saussure himself.

Anderson traces the second source for this position to the impact of the field of information theory on early generative phonology, under the influence of Jakobson and Halle. During the 1940s and 1950s there was a stress on the elimination of redundancy as a necessary step in identifying the information content of a message. But, argues Anderson, 'given the our current understanding of the sheer bulk and internal redundancy of the mental storage of information', principles of information theory which seemed persuasive to Jakobson in the early 1950's appear less relevant to the study of natural language today.

Anderson concludes — significantly for our purposes — that when the intellectual inheritance of this position is identified and examined, its persuasiveness is less than complete:

Since the claim that 'Saussure said this', so it must be true, and the notion that information theory dictates such a view — two underpinnings of the redundancy-free notion of phonological form — can thus be argued to be less than persuasive in present-day terms, we might well want to reevaluate our assumptions in this area

#### 4.2.1. *Reciprocal dependency*

We referred above to the problem that reciprocally dependent properties pose for a theory which attempts to omit predictable information. In view of the importance of this link in Anderson's argument, we will elaborate on this point. Anderson's examples of the problem both involve syntagmatic dependencies between segments, and even confining attention to this one type of dependency, Anderson observes that the problem is widespread. However, there are other classes of reciprocal dependencies not mentioned by Anderson — involving context-free dependencies between features in segment structure constraints, and involving dependencies between feature composition and syllable structure. Consideration of these cases underlines the pervasive nature of the relation of reciprocal dependency in natural

language, and intensifies the objection that a theory powerless to express this property is severely deficient. We illustrate each of these possibilities in the following paragraphs.

- *Inter-segmental dependencies: default values*

Although both of Anderson's examples involve segment-segment dependencies, they illustrate two sub-cases of quite different natures. The first involves alternations and the notion of 'default value'; the second involves non-alternating forms and values which are predictable by logical implication.

The first example involves Russian vowel-consonant dependencies. Most Russian vowels are either inherently [+back] (/o/, /u/, /ɑ/) or inherently [-back] (/e/). The high, non-round vowel, however, has front [i] and back [ɨ] variants, depending on whether it occurs with palatalized or non-palatalized consonants respectively. We might argue, then, that the value of [back], being predictable, should be omitted from underlying representation, yielding an archiphoneme /I/, unspecified for [back].

In addition, most Russian consonants are either inherently 'soft' (palatalized), or inherently 'hard' (non-palatalized). The velar obstruents (/k/, /g/, and /x/) however, have hard and soft variants, depending on whether they occur with [±back] vowels. Again, since palatalization is predictable, we might argue that this feature should be omitted from the underlying representations of these segments.

The problem, of course, arises when an underspecified consonant is combined with such an underspecified vowel. The vowel's place feature is dependent on the consonant's place feature, but reciprocally, the consonant's place feature is dependent on the vowel's place feature. As it turns out, in such combinations in Russian the vowel is always front and the consonant is always palatalized. As Anderson (1985:11) points out, if the vowel /I/ has what amounts to a DEFAULT [-back] specification...

...there is no problem: we need only say that (a) velars become palatals before front vowels, and (b) /i/ becomes [+back] after 'hard' consonants... If phonological elements are only specified for their nonredundant properties, however, the rule of velar palatalization cannot make reference to the frontness of a following /i/

We need to make an important terminological clarification here. Anderson uses the term ‘redundant’ indiscriminately to refer to IMPLIED properties — properties which are determined by the presence of some contextual property — and DEFAULT properties — properties which occur precisely in the *absence* of any determining contextual property. Moreover, he uses this same term, ‘redundant’, to refer to properties which are predictable but which alternate — the specification for [back] on /I/, for example — and also to refer to properties which are predictable but do *not* alternate — the specifications for [high] and [back] on /a/, for example. We can bring out the difference in the following way: the former characterize ‘underdetermined’ objects (their value for F can be either value); the latter characterize ‘predetermined’ objects (their value for F cannot but be as it is). In this thesis, we will restrict the term ‘redundant feature specification’ to mean ‘predictable, non-alternating feature-specification’.

The distinctions just drawn will be crucial in subsequent chapters, and will be formally distinguished in the theory we develop, which advocates the use of DEFAULT SPECIFICATION, PARTIAL SPECIFICATION, and STRUCTURED SPECIFICATION respectively. We will see that Anderson is representative of the field in this respect, and will argue that these three sub-types of specification have been not been adequately distinguished — conceptually or formally — in generative phonology, with deleterious consequences. The formal analogue of Anderson’s terminological conflation is the decision to encode all types of ‘redundancy’ by *omission* of the ‘redundant’ features. On the contrary, we will propose a model which encodes the three types of specification using distinct, but related, structural relations between features.

- *Inter-segmental dependencies: implicational values*

Anderson’s second case of reciprocal dependency is more straightforward, involving a biconditional implicational relationship between properties of adjacent segments. As just noted, although Anderson presents this case as of the same type as previously discussed, we observe that it involves neither defaults, nor alternation:

given a cluster of nasal plus stop in many languages, it is possible to predict the point of articulation of either from that of the other. Typically, we specify one of the properties (e.g. the articulation of the stop) phonologically, and include a rule to introduce the other (the articulation of the nasal, in this example). We must realize, however, that, from the point of

view of eliminating redundancy, the decision to eliminate one of two such interdependent properties is either completely arbitrary or at least based on ancillary principles of a somewhat ad hoc sort which are seldom made explicit or precise (p. 11)

This then, is a case of non-alternating reciprocal dependency in the phonotactics, a biconditional 'sequence structure condition'; it has an analogue in segment structure conditions, and to this we now turn.

- *Intra-segmental dependencies*

Consider the following fully specified matrix:

	A	B	C	D	E	F
h	+	+	+	-	-	-
f	+	+	-	+	-	+
g	+	-	+	+	-	+

Note that  $[\pm h]$  splits the set into two disjoint classes. Within the  $[+h]$  class, the features  $[f]$  and  $[g]$  are needed to distinguish /A/, /B/, and /C/; within the  $[-h]$  class, however, the features  $[f]$  and  $[g]$  co-vary. For this class, a segment is  $[+f]$  iff  $[+g]$ , and  $[-f]$  iff  $[-g]$ . The biconditional formulation is symptomatic of the fact that, in this environment, the features  $[f]$  and  $[g]$  are reciprocally dependent. The standard response to such a situation in generative phonology — pre-SPE, SPE and post-SPE — is to *omit* one or other of the features in this environment; however the choice as to which feature to omit is arbitrary.

Nor is this simply for a problem for a rule-based (as opposed to a condition-based) account. As Stanley observes (1967:435), a condition-based approach is subject to the same arbitrariness. The problem is with the decision to *omit* redundant specifications, irrespective of whether the omission is rule-based or constraint-based.

Moreover, this situation is extremely common in phonological systems. Consider the following canonical five-vowel matrix:

	i	e	o	u	ɑ
low	-	-	-	-	+
round	-	-	+	+	-
back	-	-	+	+	+



Here, it is the feature [ $\pm$ low] that splits the inventory into two disjoint classes; and within the [-low] class, the features [back] and [round] are reciprocally dependent. As it happens, in this case we can omit one or other of these features *throughout* the inventory, not just within the class in which they co-vary. But this does not alter the fact that the choice as to which feature to omit is arbitrary: if anything, it highlights it.

- *Segment-syllable dependencies*

The problem that mutually dependent properties raise for the principle of underspecification has again recently been noted by Mohanan (1991:306). The type of mutual dependency he cites is the following: given information about the feature composition of the segments of a string, we can predict its syllable structure; but equally, such syllable structure information allows us to predict information about the feature composition of segments. We assume that the mutual dependency of such aspects of phonological structure is uncontroversial. Mohanan concludes that the assumption that 'underlying representation may not contain the specification of predictable information' is thus untenable, since there is no optimal way of satisfying the assumption without being contradictory.

Mohanan (1991:307) observes:

The challenge that this problem raises for [the assumption of underspecification] has often been overlooked in the literature.

We might speculate that generative phonology's willingness to live with this inconsistency is due in large part to the absence of a viable alternative theory of specification which is well motivated and empirically adequate. It is such an alternative that we seek to supply in this thesis.

What, then, do we take away from Anderson's overview? First, despite the fact that the position that underlying representations are redundancy free is entrenched in modern phonology, there are outstanding problems with this position, and room for a logically coherent alternative. Second, this position was introduced as an integral part of the very first formulations of generative phonology, which attempted to provide representations 'from which the last drop of redundancy has been wrung' (Anderson 1985:325). And moreover, this attempt was a direct result of the influence of Jakobson's and Halle's concern with information theory.



#### 4.2.2. *The history of specification theory*

In broad outline, the history of 'specification theory' in generative phonology falls into three eras:

##### ***I. Early generative phonology***

In the first formulations, as exemplified in Halle (1959) lexical representations were redundancy free, and underspecified for redundant specifications. Phonological rules had direct access to these underspecified representations. Redundancy rules subsequently inserted the omitted specifications in the course of the phonological derivation.

##### ***II. Stanley and SPE***

This position was radically revised due to the arguments of Stanley (1967). Stanley showed that the treatment of redundancy in terms of extrinsically ordered feature insertion rules (morpheme structure rules) had undesirable formal consequences, and permitted the formulation of basically incoherent rule systems. Stanley suggested two important revisions: (i) redundancy statements should properly be expressed in terms of conditions, rather than rules (ii) phonological rules should not have access to underspecified representations: representations must be fully specified prior to the application of any phonological rule. Stanley's arguments are especially important for the history of generative phonology, since they were accepted completely and unconditionally by Chomsky and Halle in *The Sound Pattern of English*. Thus it is Stanley (1967) which is the source of what was to become the orthodox position on specification theory.

According to this position, while 'dictionary matrices' or lexical representations were underspecified for redundant information, they 'became' fully specified prior to the application of phonological rules. SPE draws a distinction between LEXICAL REPRESENTATION (which is underspecified) and UNDERLYING REPRESENTATION (which is fully specified): the former is the stored form; the latter is the form which is made available to the phonology. Further, there is no rule-based derivation mapping between these two representations: the relation between them is expressed in terms of conditions, rather than rules. In this model redundancy statements and phonological rules proper are kept totally disjoint.

### *III. Underspecification theory*

In recent years (since Kiparsky 1982), there has been another revolution in the treatment of redundancy. It follows from the model of SPE that, by the time a phonological rule applies, the distinction between redundant features and distinctive features has been obliterated. Without some kind of transderivational access to the earlier stages of the representation, a redundant feature and a distinctive feature present exactly the same appearance to a phonological rule. The very organization of the model predicts, then, that phonology proper is insensitive to the redundant vs distinctive status of the features it manipulates. This has been shown to be incorrect, however. Numerous cases have been discovered where phonological rules are seen to be sensitive to the distinctive status of features (Steriade 1987, Clements 1987). In the face of this evidence, however, theoreticians have not, as might be expected, developed a new theory of specification, but have basically returned to the earlier, pre-SPE approach, but this time attempting to meet the objections raised to such an approach by Stanley. In this theory of specification — so-called UNDERSPECIFICATION THEORY — phonological rules once more have access to representations which lack redundant specifications.

At first glance, this seems to be a totally revisionist position, which stands in clear opposition to that of SPE. There is a sense, however, in which underspecification theory is simply the logical consequence of interpreting the SPE theory of specification in the light of lexical phonology. In SPE, lexical entries are underspecified for redundant information; but equally in that theory, all phonological rules are 'post-lexical'. If, however, one admits the central proposal of lexical phonology, that certain phonological rules operate in the lexicon itself, then it is natural to expect some interaction between the application of those rules and the underspecified nature of lexical forms. Rather than thinking of the approach as 'delaying' application of redundancy rules until after the application of phonological rules, as in early generative phonology, underspecification theory might more properly be thought of in terms of the 'anticipation' of the lexical redundancy rules by the rules of lexical phonology.<sup>1</sup>

---

<sup>1</sup>It is no accident that the first cited source for the position of underspecification theory is Kiparsky (1982).

This move has, however, sharpened the controversy over the ‘rules or constraints’ status of redundancy statements. Stanley and SPE come out clearly for the position that redundancy statements are best stated as conditions, rather than blank-filling rules. Underspecification theorists have thus been confronted with the problem of how to formulate the interactions between redundancy statements and phonological rules proper. This problem has not been adequately dealt with in underspecification theory. Redundancy statements have been simultaneously couched in terms of blank-filling rules *and* conditions on representations: that is, researchers have (implicitly) been working with the (inconsistent) hypothesis that they are *both* rules and conditions, as the demands of the phonological derivation dictate. These issues will be amplified in the subsequent chapters, when we compare our approach in detail with the approach of underspecification theory.

#### 4.2.3. *Conclusion and Introduction*

We accept that the work in underspecification theory has established that phonologies are sensitive to the distinctive or redundant status of features. We also accept Anderson’s and Stanley’s objections to the formal technique of *underspecification* as an appropriate way to formally express this aspect of phonological properties. In the next chapter we develop an alternative approach to specification theory — structured specification — which enables us to preserve the insights of underspecification, while avoiding the problems inherent in the underspecification approach.

While we distinguish our position very clearly from all three approaches above, we also wish to stress the aspects of continuity: we will in fact borrow ideas from all three approaches. We are clearly indebted to the work in underspecification theory, for establishing the central and conditioning role of redundancy in phonological rule application. We also, as will become clear below, see our approach as the execution of Stanley’s ‘last word’ on the nature of redundancy, in his influential (1967) paper. And, perhaps most surprisingly, we find parallels between our approach and one of the theoretical devices of the earliest generative phonology, a device which was responsive to information theoretic concerns; which was severely criticized by Stanley; and which was subsequently excluded from the framework of generative phonology: the branching diagram. It is to this earliest formulation of the generative approach to redundancy that we turn in the next section.



Two final aspects of our approach are relevant to Anderson's commentary on the history of specification theory. Anderson makes the observation, in his discussion of the problematic nature of reciprocally dependent properties, that 'among twentieth-century phonologists, only the British Prosodic school have been willing to take this point seriously enough to reconsider the basis of the role played by redundancy in linguistic descriptions'. We acknowledge the Firthian dimension to the approach we develop, and in a later section, draw explicit parallels between our model of lexical organization and that proposed by Waterson (1987).

Secondly, as mentioned in passing above, Anderson makes the point that, in the field of speech science, some researchers adopt the hypothesis that 'the only linguistically significant representation of linguistic forms which speakers manipulate is one which is maximally specified down to very low levels of phonetic detail'. We take it to be a significant advantage of our approach that it provides a bridge between the apparently dichotomous attitudes to redundancy taken by generative phonology on the one hand, and speech technology on the other. We posit representations which encode the phonologically significant notions of redundancy and distinctiveness, yet we also hold that these representations are maximally specified. To anticipate, the possibility of maintaining such a position follows from the simple notion, argued for below, that distinctiveness, and its dual redundancy, are *relations* between properties. In order to resolve the apparent conflict between minimal and maximal specification, we simply need to encode redundancy relations between features as an integral part of the representation, and to sensitize phonological rules and constraints to these relations.

### **4.3. Halle's approach to redundancy**

#### *4.3.1. Introduction*

In this section we review the approach to phonological redundancy advocated in Halle (1959). We have two objectives: firstly, to review explicitly the motivations for underspecification in generative phonology by studying the first phonological model informed by information theoretic considerations; and secondly, to review the function and structure of BRANCHING DIAGRAMS in early generative phonology. We will argue subsequently that the utilization of a hierarchy related in a number of ways to that of the branching diagram obviates the need for underspecification.

#### 4.3.2. The Minimality Condition

Halle's *The Sound Pattern of Russian: a Linguistic and Acoustical Investigation* is best known for its introductory section, 'A Theory of Phonology', which sets out SIX FORMAL CONDITIONS which phonological descriptions must satisfy. It includes the celebrated rejection of the bi-uniqueness requirement (Halle's Condition (3a)) as 'an unwarranted complication which has no place in a scientific description of language' (p. 24). In this thesis we are primarily concerned with Halle's Condition (5), the first part of which runs as follows:

Condition (5): In phonological representations the number of specified features is consistently reduced to a minimum... (p. 29)

In addition to the six conditions, Halle (1959:32) also contains the first formulation of what Stanley was later to call the DISTINCTNESS CONDITION:

We define the following order relation between segment types: Segment-type {A} will be said to be *different from* segment-type {B}, if and only if at least one feature which is phonemic in both, has a different value in {A} than in {B}; i.e. plus in the former and minus in the latter, or vice versa.

Although Halle uses the term 'different from', we will refer to this relation as 'distinct from', following standard practice.

Many of the features of Halle's model will be familiar to anyone acquainted with generative phonological descriptions, but there are less well-known aspects of the analysis which will be crucial to the argumentation below.

After introducing the familiar concept of representations couched in terms of distinctive features, Halle distinguishes between two types of feature: *phonemic* features, which serve to distinguish one morpheme from another and *non-phonemic* features, which are distributed in accordance with 'general rules of the language and hence cannot serve to distinguish one morpheme from another' (p. 29). This distinction is familiar enough,<sup>2</sup> as is the suggested way in which the distinction should be encoded in the grammar: the elimination of redundant information from underlying representations. It is instructive, however, to consider the rationale provided for this particular way of encoding the distinction. Halle (1959:29) suggests that phonemic

---

<sup>2</sup>Familiar, but, as we will see subsequently, incorrect.

features represent *instructions for selecting a particular morpheme from a list*, and that due to the rapidity of speech, these instructions need to be executed as *efficiently* as possible:

If the grammar presented here is to be taken as a realistic picture of the functioning of a language, then the instructions for selecting a particular morpheme represent a conscious effort by the speaker, in contradistinction to the operation of various obligatory rules of the language, which the speaker follows automatically when speaking a particular language. Since we speak at a rapid rate — perhaps at a rate requiring the specification of as many as 30 segments per second — it is reasonable to assume that all languages are so designed that the number of features that must be specified in selecting individual morphemes is consistently kept at a minimum. This assumption is embodied in the following requirement:

Condition (5): In phonological representations the number of specified features is consistently reduced to a minimum compatible with satisfying Conditions (3) and (4)

Note that the omission of redundant information is *not unconditional*, as the final part of Condition (5) indicates; we return to this point in a moment. But even ignoring this for the moment, we see that the rationale for omission comes down to this: we require an efficient encoding. It is this notion of *SELECTIONAL EFFICIENCY* that is primary, and taken from information theory; Halle's solution — the adoption of a principle which we might call *REPRESENTATIONAL PARSIMONY*, as embodied in Condition (5) — is a particular execution of this leading idea. Halle's argument is that the selection operation might reasonably be expected to be as efficient as possible: parsimonious specification is one way to achieve this, but obviously not the only way. Rather than simply omitting certain information, a reasonable alternative is simply to provide a *structured organization* of information specifically designed to allow the selection procedure to function efficiently. There is absolutely no *a priori* reason that any information at all need be suppressed to achieve this goal. As we will see subsequently, this is one way to characterize the approach we advocate. Further, if an important determinant of the nature of specification is taken to be its role in the efficient identification of morphemes, there are compelling reasons to think that redundant information is crucial to such an enterprise.



As Anderson notes, the adoption of minimal specification in generative phonology owes much to the development of the field of information theory, which stressed the elimination of redundancy as a necessary step in identifying the information content of a message. But Halle's approach fails to express an equally important principle of information theory, a principle put succinctly by Hockett (1987) 'the only defense against noise is redundancy'. This principle was, indeed, well known to Halle and Jakobson, as is evident from the following observation (Cherry, Halle and Jakobson 1953: 39):

The term 'redundancy' should not be taken to imply wastefulness; it is a property of speech, and in fact of every system of communication, which serves a most useful purpose. In particular, it helps the hearer to resolve uncertainties introduced by distortion of the signal or by disturbing noises.<sup>3</sup>

There is an apparent paradox here: it is explicitly claimed that redundant information improves the efficiency of identification of morphemes in the face of noise, and yet in Halle's model redundant information is omitted from phonological representations for the precise reason that it improves the efficiency of lexical access. In effect, the representational theory of generative phonology drives a wedge between the procedures of 'physical identification' and 'mental identification', and gives precedence to the latter. In view of the formal problems that such a view encounters, it appears that this hard distinction is somewhat artificial. It has also led to the parting of the ways between generative phonology and speech technology mentioned earlier, since it artificially divorces the treatment of recognition in the speech technology sense, from what might be termed 'phonological recognition'. On the contrary, we will propose a theory of representation which is efficient, but which gives equal prominence to distinctive and redundant information.

Returning to the exposition of Halle's model, we have so far a familiar picture, in which we distinguish phonemic and non-phonemic features, and systematically omit non-phonemic, predictable information from

---

<sup>3</sup>Like the term 'generate', the term 'redundant' has a common-sense meaning that tends to undermine its technical sense. We may be told and understand that 'generates' is simply a relation, but theorists have constantly been influenced by the underlying procedural metaphor; similarly, we must remember that redundant information (in the technical sense) is *not* redundant (in the colloquial sense). The colloquial connotations have not perhaps been totally dormant in the treatment of redundancy in generative phonology.

representations. Such non-phonemic features are subsequently re-introduced by rules, termed by Halle *F rules* (p. 37).

#### 4.3.3. *Caveat I: 'Condition (4)'*

However, Halle notes immediately that this position is too strong, and that certain information, despite being predictable, must nevertheless be present in the phonological representation, since this information is required by 'transformational' (morphological rules):

It may [...] seem most advantageous to put all rules governing the distribution of non-phonemic features after the transformational rules. There are, however, reasons which make it desirable to apply some of the F rules before the transformations, even if that entails complications [...] In Russian and in many other languages — though not perhaps universally — there are transformational rules, in particular rules of inflection and derivation, which require for their proper operation that certain features be specified in the representation regardless of whether or not these features are phonemic.

In fact, this is the reason Halle includes the caveat concerning 'compatibility conditions' at the end of his Condition (5), one of which, Condition (4) is expressed as follows:<sup>4</sup>

Condition (4): The phonological description must be appropriately integrated into the grammar of the language. Particularly, in selecting phonological representations of individual morphemes, these must be chosen so as to yield simple statements of all grammatical operations — like inflection and derivation — in which they may be involved.

It is necessary therefore, Halle argues, that *certain* rule-based, non-phonemic information be present in the phonological representation and supplied to other components of the grammar, for example, morphological rules. Rather than simply *retaining* such information in lexical entries for morphemes, however, Halle chooses first to omit it, and then to re-introduce it by rules prior to the application of any transformational rules which require that information. Halle notes that this partitions the F rules into two parts:

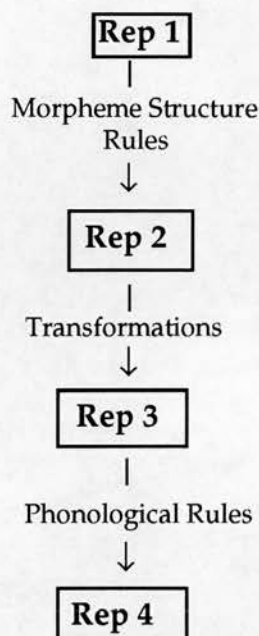
one part, which shall henceforth be called the *morpheme structure* or *MS rules*, must be applied before the transformations, and the

---

<sup>4</sup>Although the principle embodied in Condition (5) is frequently cited, its accompanying caveats are not so widely published.

other part, to be called the *phonological* or *P rules*, must be applied after the transformations. (p. 38)

We thus have the following organization:



Thus, for Halle, some redundancy rules are (those applying before the phonological rules) morpheme structure rules, but others are simply phonological rules. Stanley was to reverse this decision, and argue that all redundancy rules are morpheme structure rules, in the sense that all redundancy rules should apply in a set, prior to the application of any phonological rule proper.

#### 4.3.4. *Caveat II: The Distinctness Condition*

This, then, is the first sense in which Condition (5) — the omission of redundant or predictable information — is too strong, and is constrained by Condition (4). There is another condition which constrains minimality — the distinctness condition. Halle notes rightly that the omission of *all* predictable information results in representations which are NON-DISTINCT, in the sense defined above. Accordingly, certain F rules will need to be applied in order to insert sufficient information to render representations distinct. We illustrate these two 'degrees' of specification with an example used by Halle himself. Consider the following fully specified feature matrix:

*Full-specification:*

	t	s	c	n
strident	–	+	+	–
nasal	–	–	–	+
continuant	–	+	–	–

Some of the properties in these representations are predictable, since they follow from general implicational relationships that hold between features for the entire inventory: all stridents are oral, for example, so the fact that /s/ is strident is predictable. The complete set of such implications are as follows:

[–strident]    ⊃ [–continuant]

[+continuant] ⊃ [+strident]

[+strident]    ⊃ [–nasal]

[+nasal]        ⊃ [–strident]

Omitting predictable information results in the following matrix:

*Minimal specification:*

	t	s	c	n
strident	–		+	
nasal	–			+
continuant		+	–	

However, certain pairs of segments are now non-distinct — /t/ and /s/ for example — for example. Thus the representations here are minimally specified in accordance with Condition (5), but are insufficiently specified to meet the distinctness condition. Accordingly, certain F rules must be applied, to render each segment distinct from every other:

*Distinctive specification*

	t	s	c	n
strident	–	+	+	–
nasal	–			+
continuant		+	–	

We have seen, then, that certain predictable information must be included in phonological representations, but we have also seen that Condition (5) requires that information be kept to a minimum. How can these best be reconciled? How can we ensure that representations be minimally specified



yet distinct. It is for precisely this role that Halle introduces the notion of the BRANCHING DIAGRAM.

#### 4.3.5. *The role of the branching diagram*

The term used for representations with the level of specification seen in this last (distinctively specified) matrix is FULLY-SPECIFIED MORPHONEME. When introducing the term (Halle 1959:34) comments:

Like all segment-types occurring in phonological representations, fully specified morphonemes are subject to Condition (5), which requires that the number of specified features be kept at a minimum.

(Note a terminological point here: a 'fully-specified' morphoneme is *un*-specified for a number of features — indeed, is minimally specified consistent with the distinctness condition. The interpretation is thus 'fully-specified: for a morphoneme'. We will examine less-than-fully specified morphonemes subsequently in due course.) Halle continues:

It can be shown that imposing this condition [Condition (5)] on a set of fully specified morphonemes is tantamount to requiring that the set of fully specified morphonemes be mappable into a branching diagram

Thus certain predictable information must be inserted by MS Rules into minimally specified, non-redundant representations, in order to satisfy both Condition (4) and the Distinctness Condition. The branching diagram is employed in order that the redundant information included be kept to a minimum in accordance with Condition (5). Halle (1959:38) summarizes as follows:

the set of segment-types occurring after the application of the MS rules is defined by all possible paths through the branching diagram which begin at its initial node [...] this places a restriction on the features that can remain unspecified: certain non-phonemic [read 'predictable' MB] features must be specified at this point. The result, however, is precisely the one desired, for [...] unless some such limitation on the occurrence of unspecified features is imposed at this point, it will not be possible to apply properly the (transformational) rules of inflection and derivation

This, then, is the primary role of branching diagrams in Halle's model, the role of ensuring that fully-specified morphonemes be MINIMALLY DISTINCT. He also, however, identifies a secondary role. As Halle (1959: 34) observes:

Mapping a matrix into a branching diagram is thus equivalent to establishing a hierarchy among the features

We will examine the way this hierarchy arises in a moment, but here we simply note the significance Halle (1959:34) attaches to it:

The hierarchy of features seems to provide an explanation for the intuition that not all features are equally central to a given phonological system; e.g., the distinction between vowels and consonants is more fundamental to phonological systems than the distinction between nasal and oral vowels, or voiced and voiceless consonants.

#### 4.3.6. *The form of branching diagrams*

Having examined some of the motivations for the inclusion of branching diagrams in the theory of phonology, we now turn to an examination of their structure. Recall that the primary motivation is to ensure that F rules apply in such a way as to convert a minimally specified matrix into a distinctively specified matrix by including the minimum amount of information. If a matrix can be mapped into a branching diagram, then its degree of specification is, it is claimed, both minimal and distinctive.

A branching diagram is a binary-branching tree, with a feature assigned to each node, and the two branches representing the two values the feature may assume (plus or minus). Each node of the tree is also associated with a class of segments, the complete class being associated with the top (root) node. In each local tree<sup>5</sup> the class at the mother is partitioned into two subclasses, one at each daughter, according to the binary classification of the relevant feature. An important point to note is that the same feature may be used as the basis of classification for different local trees.

The mapping between the matrix and a tree is defined in the following way: a feature-value will be specified for a segment in the matrix if and only if it is used in the classification of the segment (at any level) in the branching diagram. In other words, the mapping ensures any segment A in the matrix is explicitly coded as disjoint from any other segment B, either directly, or as belonging to a class which is disjoint from the class to which B belongs. In

---

<sup>5</sup>A tree of depth one, consisting of a mother and the two daughters.



addition, since this is the only criterion of specification, it attempts to ensure that the representations are coded as disjoint using the least amount of information, just the information required to achieve distinctness.

There are two immediate problems with this model. Firstly, as is noted by Halle, and has been criticized by Stanley (1967) there is an inherent indeterminacy in the classification. Secondly — and this point has not been sufficiently appreciated — as our exposition above makes clear, the classification that forms the basis of the branching diagram has nothing whatsoever to do with redundancy or predictability: it is a totally orthogonal classification. This seriously compromises Halle's model. On the one hand, it is argued that specifications should be omitted by virtue of being predictable. On the other hand, it is argued that specifications be omitted because they are non-distinctive. The pattern of specification that results from the constraints of the branching diagram actually obscures the pattern of redundancy — the implicational relationships between predictable and non-predictable information — that obtains in the system. We now examine each of these points in more detail.

#### 4.3.7. *Problem 1: Indeterminacy*

Halle (1959:36) points out that:

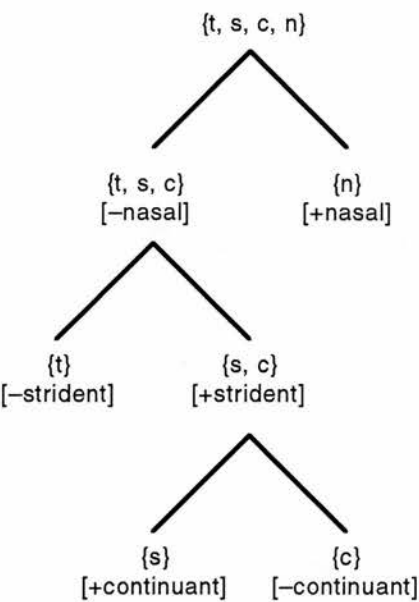
The freedom in ordering feature-questions may result in several branching diagrams compatible with the above requirements.

He illustrates his point with a partial system of four consonants 'quite similar to that of Russian'. The fully specified matrix was given above. Here is a distinctively specified matrix:

	t	s	c	n
1. nasal	—	—	—	+
2. strident	—	+	+	
3. continuant		+	—	

The feature [nasal] partitions the inventory into two classes. One class of this partition is singleton — /n/ is isolated as [+nas] — and the remaining [—nasal] sounds may then be partitioned with reference to the feature [strident], which isolates /t/. Finally, the feature [continuant] is employed to distinguish /s/ from /c/. Thus the feature [nasal] is used to subclassify the complete set of segments, whereas the feature [continuant] is used to

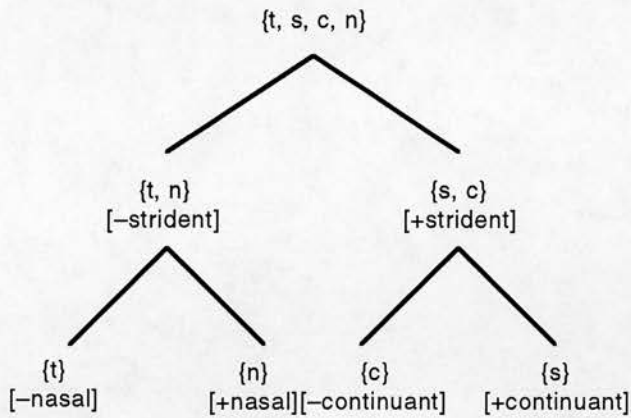
subclassify a much smaller set. The branching diagram associated with the above matrix may be represented as follows (Halle 1959:36):



However, there is an alternative way of ‘ordering the feature-questions’ — rather than starting with [nasal], we can start with [strident], for example. If we do, we will obtain a differently specified matrix for the same set of sounds and the same set of features:

	t	s	c	n
1. strident	–	+	+	–
2. nasal	–			+
3. continuant		+	–	

Here, it is the feature [strident] that is taken as primary (its ‘feature-question’ is ordered first, in Halle’s words), and which thus occurs at the root of the hierarchy, with nasal ordered below:



Halle (1959:36) observes that:

In such cases the choice may be dictated by Condition (5) which, in terms of the branching diagram, means that preference be given to the more symmetrical diagram

In this particular case, Halle (1959:36) notes that the second ordering is more economical 'since it yields the greater number of zeros — a fact which is reflected in the greater symmetry of its associated branching diagram'. We thus see that the use of the branching diagram does not *itself* ensure that the minimality condition is met; rather it produces alternative solutions, and the choice between them is itself determined with reference to the minimality condition.

However, the appeal to economy is not straightforward. Which are we to prefer: a diagram with high-level symmetry dominating low-level asymmetries, or a high-level asymmetry which captures low-level symmetries? In practice, any diagram will necessarily be a trade-off between these two tendencies, contributing to massive indeterminacy in the construction of the 'best' diagram for any reasonably complete set of features. Moreover, the only way to choose between potential diagrams is to generate and compare them, but given  $n$  features, there are  $n!$  orderings of those features, and thus  $n!$  potential diagrams. Halle, in fact, includes a branching diagram corresponding to a matrix employing 11 features. Thus the minimality condition is required to choose among  $11! = 39,916,800$  potential diagrams.

#### 4.3.8. Problem 2: The masking of redundancy relations

Recall the redundancy free, or minimally specified version of the example matrix:

	t	s	c	n
strident	–		+	
nasal	–			+
continuant		+	–	

As Halle (1959:35) observes, a matrix of this general form cannot be converted into a branching diagram because ‘it lacks a feature without a zero’. Thus there is no way to partition the set of sounds in terms of a yes/no answer to any feature question; and thus no way to ‘root’ the feature hierarchy.

Accordingly, we are compelled to *include* redundant specifications for some feature or other, in order to provide a mapping to the branching diagram, in order to satisfy the distinctness condition. Thus the formalism in fact obscures the demonstrable pattern of redundancy that holds in the system. And worse, this problem is not simply a peculiarity of the root of the hierarchy, a quirk of the ‘first’ feature question. It may be repeated at any point down the tree, since at every choice-point, the diagram requires that some feature be the means of unambiguously *partitioning* all segments under that node into two classes, irrespective of its predictable or unpredictable status. As Halle (1959: 34) writes:

The possibility of mapping a distinctive feature matrix into a branching diagram hinges upon the existence in the matrix of a least one feature for which there are no zeros. This feature, which must be assigned to the first node, subdivides the segment-types into two classes. The next two nodes must be assigned to features which have no zeros for any other segments in the two sub-classes [...] The same procedure must be again be possible with regard to the segment-types in each of the four-subclasses established by the former features; etc.

Since the construction of the tree is a recursive procedure, a problem for the procedure is a recursive problem.

#### 4.3.9. Summary

There are thus serious problems associated with the notion of branching diagrams; and in subsequent chapters we will consider even more. However, we will nevertheless argue that this notion, suitably modified, provides a

deep and unifying insight into the interrelationships between parts of generative phonology. The notion that there is a single, hierarchical organization which simultaneously (i) is concerned with the issues of redundancy and distinctiveness (ii) conditions what information is made available to morphological and phonological rules (iii) ranks all features according to some notion of their 'centrality' to the phonological system (iv) formally expresses the notion of the archiphoneme, is one which we completely support. In subsequent chapters we will argue that all these issues are indeed interrelated, and further that there is a single hierarchical organization that gives formal expression to them all. Thus while we take exception to virtually all specific aspects of Halle's presentation of branching diagrams, we will also argue that, rightly formulated, it is a powerful, explanatory and necessary component of generative phonology.

#### **4.4. Stanley's approach to redundancy**

##### *4.4.1. The Stanley model*

As we have seen, Halle's approach to redundancy was criticized by Stanley (1967), who made a numerous modifications to the model:

- He explicitly criticized the role played by branching diagrams in the phonology.
- He rejected the distinctness condition.
- He required that all redundancy rules apply together, before the operation of any phonological rule.
- He argued that redundancy rules be unordered, and that they be prevented from altering feature values.

All these considerations were then combined in the resulting model, which argued that all of the above follow from the decision to treat redundancy rules as *conditions* on lexical representations, rather than as rules mapping from one level of representation to another. All of Stanley's arguments were accepted by Chomsky and Halle (1968) who adopted Stanley's constraint-based approach.



We will consider Stanley's criticisms of branching diagrams in detail in the next chapter; and we will examine his argument against the interleaving of redundancy rules with phonological rules in our discussion of underspecification theory in Chapter 9. We note, however, that Stanley's final formulation, even though a conditional formulation, could be, and was, used to omit specifications from underlying representations. Stanley (1967:421) however, explicitly rejected the distinctness condition, replacing it with the following criterion for determining when two dictionary matrices are different enough to be kept apart:

two dictionary matrices are properly distinguishable just in case the MS rules render them distinct, where 'distinct' has [its] formal meaning

(Recall that, for Stanley, MS rules are unordered and monotonic). The point is that, for Stanley, the specifications that underspecified representations omit are just the redundant (predictable) specifications; and equally, the specifications that do appear in underspecified representations are just the unpredictable specifications. There is no mention of distinctive specifications, and there is no level of specification which contains a minimum amount of 'distinctive' information, in addition to the properly unpredictable information. There is no need to include some minimal amount of distinctive information. Thus the underspecified representations will be MINIMALLY specified in the sense we defined above. To put it another way, for Stanley, (as for us) the notion of a 'minimally distinctive' level of representation, mixing as it does criteria of distinctiveness or contrastiveness on the one hand, and predictability on the other, is an incoherent level.

#### 4.4.2. *Stanley's alternative proposal*

The model described in the previous section is well-known, forming as it did the basis of the SPE approach to specification. Despite the differences between this model and the model it replaced, there is a central common feature: both models choose to formalize the notion 'redundant' feature value by leaving that value blank, and then determining the value by some 'finite procedure' in Stanley's phrase. Stanley also observes that this notion is also common to an approach couched in terms of MS rules or MS conditions. We now focus on the closing section of Stanley's (1967) paper. This section appears as an appendix, as it were, to the main argument, and contains a radical proposal for the treatment of redundancy in phonology, one which was not

incorporated into either Stanley's own model or the model of SPE. This is the proposal to abandon blanks entirely.

In a section titled 'The Role of Blanks' Stanley (1967:433–4) entertains the following proposal:

Suppose we proceed differently and abandon the idea of having a dictionary matrix for each morpheme which is less fully specified than the systematic phonemic matrix. In fact, let us abandon (but for the moment only) the whole notion of dictionary matrix, and talk only about fully specified systematic phonemic matrices, regarding these both as the matrices to be listed in the dictionary and as the matrices which enter the P rules. Still, let us retain in each grammar the set of MS conditions that state the constraints [...] that exist on the morphemes of the language.

Stanley notes that such an approach has an important advantage over what was to become the standard model. For it turns out that, even using conditions instead of rules, even allowing phonological rules only to access fully specified representations, the problem of reciprocal dependency still remains:

if, in some environment E, the value [+f] implies the value [+g] and the value [+g] implies the value [+f], then it would be arbitrary which value we actually choose to indicate in the dictionary [...]

As Stanley (1967:435) notes, this problem is orthogonal to the question whether redundancy be stated in terms of rules or conditions, but stems from the decision to *omit* redundant information, however this be achieved:

The arbitrariness stems from the decision to commit ourselves as to whether f is predictable from g or g from f, this decision being involved in any solution where redundant feature values are left blank in dictionary matrices. Actually, however, this decision is not well motivated, since the correct statement may simply be that the values of f and g are interrelated

Stanley (1967:435) continues with an observation that might stand as an epigraph to the approach to specification we propose in this thesis:

In fact, we can profitably draw a general conclusion regarding redundancy from this discussion. That is, to say that a certain fully specified matrix is highly redundant in some language is actually to say that many of its feature values are interrelated in ways determined by the constraints of the language [...] Once these constraints have been stated, it is true they may be utilized

in giving dictionary representations their most economical form; but this is a secondary fact, and these redundancy-free representations play no real role in a theory of redundancy.

This is (literally) Stanley's last word on the matter of redundancy, and it provides a succinct formulation of the theory of specification we will develop.

#### 4.5. The proposal: structured specification

In the approach to specification theory proposed in this thesis, redundancy relationships are taken to be primary, and are coded directly into phonological representations. In this model, like Halle's, features form a hierarchy. Unlike branching diagrams, however, our hierarchy is based on the redundancy relationships themselves, rather than the requirement of establishing non-distinct representations. In this REDUNDANCY HIERARCHY, *all* features are represented, and the redundant vs non-redundant status is coded by the interrelationships between the features. Obviously this approach does not encounter the problem (if, indeed, it be considered a problem) of non-distinct representations, since no information is omitted at any point. We employ only *one level* of representation, rather than an underspecified level and a fully-specified level. Naturally, then, there is no mapping between them, and thus no redundancy rules. But nor is there a (condition-based) PROCESS OF SELECTION mapping between two levels, as suggested by Stanley. Our approach thus constitutes a non-derivational approach to specification. The single level of representation we employ is maximally specified — like the representations used in speech sciences — but also encodes the phonologically salient distinction between redundant and non-redundant information. Furthermore, our model provides an elegant and exact execution of the notion that the statement of MS conditions itself constitutes the statement of redundancy. In our model the statement of MS conditions for the language, and the structure of redundancy exhibited by individual representations, are one and the same. Representations are stored as terminal nodes in a network — an inheritance hierarchy — the very structure of which expresses the segment structure and sequence structure constraints of the language.

In subsequent chapters, we present a formal model of redundancy which gives direct expression to Stanley's final insight — that the essence of redundancy is the interrelationship of feature values as determined by the

constraints of the language. We conclude that redundancy free representation, and thus Halle's Condition (5), has no place in a theory of phonology.

## 5. Structured specification

---

### 5.1. Introduction

This chapter presents a principled method of coding redundancy relations directly into phonological representations; an approach which we term **STRUCTURED SPECIFICATION**. The approach draws on a number of related ideas from the fields of unification-based lexicons, knowledge representation, and the algebraic theory of lattices. However, we take as our starting point the so-called branching diagrams of early generative phonology, which should be reinterpreted, we argue, as a graphic representation of the relative redundancy of features: a **REDUNDANCY HIERARCHY**.

As we have seen, there are a number of problems with such diagrams. In this chapter we argue that these objections can be overcome by a simple technical revision: rather than representing features in a tree-structured hierarchy, we represent them in a **LATTICE**. In the lattice, dominance is read as logical implication, so that, for example, the fact that [+coronal] dominates [+anterior], is equivalent to the implication or 'redundancy rule':

$$[+anterior] \rightarrow [+coronal]$$

The lattice as a whole thus provides a graphic representation of the entire set of redundancy rules. Moreover, since the lattice is based on a *partial* ordering, incomparable elements are properly left unordered, avoiding arbitrary rankings of features, and yielding a unique result for a given feature set.

This simple move has a number of implications for the theory of phonological representation and rule application:

First, it resolves the serious and long standing problem of the reciprocal dependency of phonological features, in all its forms.

Second, there is a natural interpretation of the lattice in terms of an **INFORMATION ORDERING**, with more general terms (applying to larger sets of sounds) at the top, and more specific terms (applying to smaller sets) at the bottom. The lattice representation thus reveals a connection between



the notion of DISTINCTIVENESS and the ELSEWHERE CONDITION: 'distinctive' features are 'most specific'. It follows that if distinctiveness is coded into representations in terms of specificity, then rules (or constraints) which require that a segment bear a feature distinctively will have this specificity coded into their structural descriptions, and will thus be ordered by the elsewhere condition.

Third, it provides us with a straightforward way of coding the redundancy relations among the properties of a segment into the representation of the segment itself: a segment is simply a sub-lattice of the lattice of features. Each segment lattice will have one or more 'lowest features', in a sense which will be made formally precise. These are the features whose presence is not implied by any others: in other words, the features which are distinctive for that segment.

Fourth, in addition to providing a novel approach to segment structure constraints, the approach extends naturally to sequence structure constraints. Moreover, once these are incorporated, the result is a lexicon exhaustively structured as a single lattice, a storage structure which has implications for recognition, acquisition, and lexical access. The resulting model has the interesting property that phonotactic constraints are not stated in isolation from lexical entries, but are *immanent in the structure* of storage of the lexical entries. Thus the 'phonotactic constraints' of a language are expressed through the *pattern of storage* of lexical representations.

Fifth, there is a natural extension of the approach to default properties, which also exhibit a lattice structure based on subsumption: default or unmarked properties subsume exceptional or marked properties. This allows us to apply ideas developed in the field of knowledge representation concerning NONMONOTONIC INHERITANCE; and conversely, reveals that phonological systems provide a fertile field for the exploration of empirical questions concerning the behaviour of complex systems with defaults — information which is being actively sought by theorists in the field of knowledge representation.

Sixth, it provides a straightforward formalization of the notion archiphoneme, and permits us to distinguish PARTIALLY SPECIFIED from REDUNDANTLY

SPECIFIED segments: a distinction which current phonological theory is unable to make.

Seventh, in an analogous way, it provides a means of making a principled distinction between what Hualde (1991) terms UNSPECIFIED and UNMARKED segments: a distinction which current theory is also unable to make.

Eighth, it resolves serious formal and conceptual inconsistencies which result from combining underspecification theory and the theory of FEATURE GEOMETRY, both as regards subclassification of features, and also the treatment of monovalent features.

Ninth, a consequence of the approach is that the question of whether a particular rule is sensitive to the distinctive vs redundant status of features or not must be regarded as a parameter of the rule itself, rather than a function of the degree of specification in the representation. This leads to a non-derivational approach to specification theory, an approach which makes different empirical predictions from those of current underspecification theory: predictions which we show are correct.

Tenth, the underspecification approach to redundancy makes a number of specific assumptions regarding rule-application which are incompatible with a constraint-based approach to phonology. Structured specification, on the other hand, is fully compatible with a constraint-based approach, and thus constitutes the first appropriate theory of specification for this emerging paradigm.

In this chapter, we introduce the basic principles of the theory of structured specification, reserving discussion of its implications for subsequent chapters.

## **5.2. Structured specification**

### *5.2.1. Introduction*

We now consider an alternative way of deriving a hierarchy among phonological features based on the notion of their RELATIVE REDUNDANCY. This notion is partially expressed by the branching diagrams of Halle (1959). For Halle, to use the example discussed in the previous chapter, the feature [nasal] is relevant to the distinctive classification of every segment, whereas

the feature [continuant] is relevant only to the distinctive classification of a small subset of segments. This means that for segments outside this small subset the value of the feature [continuant] will be redundant. The ordering of features in a branching diagram thus in part corresponds to an ordering from 'least redundant' (applicable to the greatest number of segments) at the top to 'most redundant' (applicable to the smallest number) at the leaves.

### 5.2.2. *Lattice Theory*

We will formalize this concept of a REDUNDANCY ORDERING by employing concepts from the general algebraic treatment of order. We will begin with the examination of segments and extend the analysis to syllables and morphemes in subsequent sections. We proceed as follows. Our starting point is the lattice of all logically possible representations (feature matrices) supported by a given set of features. We then show how the set of actual representations (segments) in a given system can be derived by 'collapsing' this lattice of possibilities in systematic ways. Next, we perform a REDUCTION (an operation which will be defined below) of the lattice labelling, distilling the information expressed in a straightforward manner. No information is lost in this reduction: the resulting lattice is simply a maximally efficient encoding of the actual representations manipulated by the phonology. However, this lattice is also straightforwardly interpreted as a direct encoding of the redundancy relations which hold between the various features of the phonology. Thus what is expressed in other approaches by a set of redundancy rules is here expressed by the very structure of the (reduced) lattice of representations. We adopt this exposition in order to provide a concrete demonstration of the notion that it is possible to move directly from representations to redundancy statements without recourse to rules of any sort. The exposition also amounts to a demonstration of Stanley's claim that the statement of the redundancy of a feature matrix actually amounts to a statement of the interrelationships between its features.

First consider a simple vowel system in which there is no redundancy to be expressed. Quileute (Maddieson 1984:379) has a four-vowel system which can be exhaustively characterized by the features [high] and [back]:

	i	a	ɑ	u
high	+	–	–	+
back	–	–	+	+

In this language, all the logically possible combinations of the two features [high] and [back] are attested. There are no gaps in the system, and thus no combination of features is ruled out. Conversely, no feature specification can be predicted on the basis of any other: there is no redundancy here since every feature 'counts'. The two features [high] and [back] are thus sufficient to describe the four segments of the inventory, but of course when used in rules the same features may be employed to describe *classes* of segments which either undergo or condition phonological processes. The total number of descriptions these features make possible, together with the classes denoted by each, are as follows:

(1)

$$\begin{aligned}
 [+hi] &= \{i, u\} \\
 [-hi] &= \{a, \text{a}\} \\
 [+ba] &= \{u, \text{a}\} \\
 [-ba] &= \{i, a\} \\
 \begin{bmatrix} +hi \\ +ba \end{bmatrix} &= \{u\} \\
 \begin{bmatrix} +hi \\ -ba \end{bmatrix} &= \{i\} \\
 \begin{bmatrix} -hi \\ +ba \end{bmatrix} &= \{\text{a}\} \\
 \begin{bmatrix} -hi \\ -ba \end{bmatrix} &= \{a\}
 \end{aligned}$$

It is essential to what follows to note that these descriptions form a partially ordered set, by virtue of the inclusion ordering on their denotations. A feature-matrix of the form traditionally used in phonology specifies a set of sounds by listing the properties that characterize the set. In general, the fewer the number of features, the greater the number of segments which satisfy the description. Conversely, the larger the number of features, the smaller the number of segments satisfying the description — until in the limiting case, we reach a single segment.<sup>1</sup> We call such a description, one which uniquely identifies some object, a COMPLETE description, and any description which

---

<sup>1</sup> There is in fact an important class of exceptions to this generalization, which will be the focus of discussion below.



falls short of this, which denotes a class of objects, a PARTIAL description. As we have just seen, it is possible to list all the possible sets of segments which our properties let us describe, from the most general to the most specific, and it is clear that in many cases one set will be a member of — included in — another, just as the set  $\{u\}$  above is included in the set  $\{i, u\}$ . It follows that we can use this property, the property of set inclusion, to *order* the sets of sounds, and by implication the descriptions associated with each set. The arithmetic ordering 'greater than' is in fact a special case of the more general algebraic notion of order. We are in effect saying that the set  $\{i, u\}$  is 'greater than' the set  $\{u\}$ , in the intuitively obvious way: it is simply that the ordering relation we are employing here is set-inclusion rather than arithmetic magnitude. From the point of view of descriptions, we are saying that some descriptions are 'more general' than others, while providing an exact definition of generality in terms of the sets of sounds denoted by the descriptions.

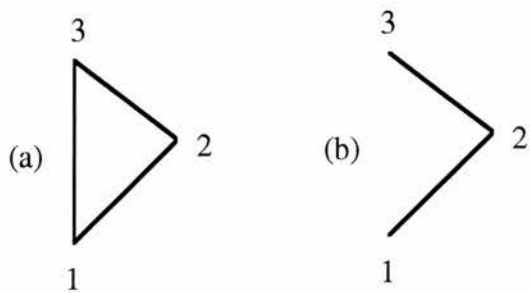
Of course, not every pair of descriptions will be related by the subset relation: some descriptions may refer to totally different sets. In terms of the subset relation, the set  $\{i, u\}$  and the set  $\{a\}$  are not related at all. In mathematical terms, we call such sets, and their associated descriptions, INCOMPARABLE, and it is for this reason that we say the descriptions form a *partially* ordered set.

One of the most useful and attractive features of ordered sets is that they may be 'drawn'. The elements of the set are arranged with the 'greater' at the top and the 'lesser' at the bottom. If there is an ordering relation between two elements, the greater element is positioned higher than the lesser and they are connected by a line; if the elements are incomparable, they are positioned on the same level, and not connected. By definition, an ordering relationship of the sort we are considering here is transitive: if  $A$  is greater than  $B$ , and  $B$  is greater than  $C$ , then  $A$  is greater than  $C$  — this is just part of what we mean by order. It follows that, when drawing an ordered set, we can suppress many of the lines and not lose any information: the set  $\{1,2,3\}$ , for example, is drawn as in (2b) rather than (2a). (In exactly the same way, and for the same reasons,



the relation of 'dominance' can easily be read off a syntactic tree, which only explicitly diagrams 'immediate dominance').<sup>2</sup>

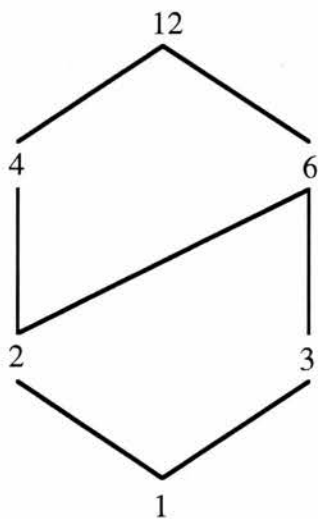
(2)



The relation of immediate-dominance in a lattice is known as the COVERING relation; thus in the diagram above 3 covers 2, 2 covers 1.

The set pictured in (2b) represents a completely ordered set. The following example represents a partially ordered set: the set of integer divisors of 12, ordered by the relation 'factor of'. Note that (i) 4 and 6 are factors of 12, but neither is a factor of the other, thus 4 and 6 are incomparable (ii) 2 is a factor of both 4 and 6, represented by a direct connecting line, but also of 12, represented indirectly via transitivity:

(3)

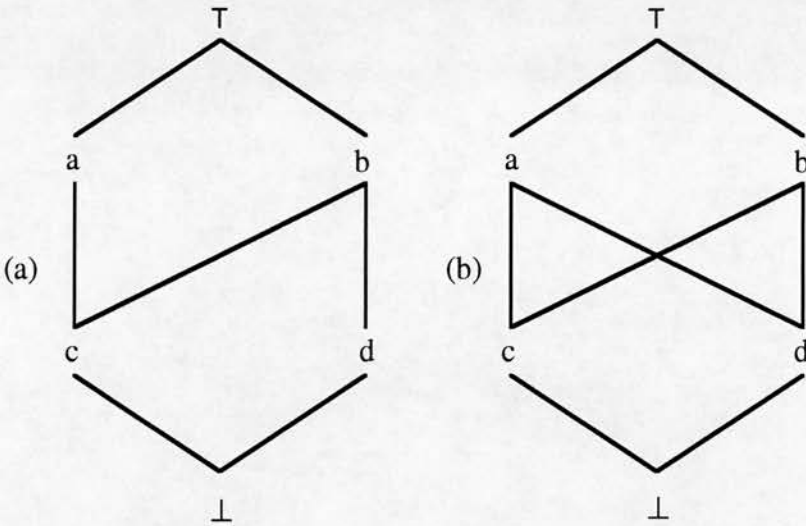


<sup>2</sup>And, for that matter, the relation 'ancestor of' can be read off a genealogical tree, which only explicitly diagrams 'immediate ancestor'. Such trees, in fact, constitute one of the earliest pictorial representations of an ordered set (Birkhoff 1948:6).

In fact, the partial order in (3) has the requisite properties to qualify as a LATTICE. Note first that it has a **unique greatest element**, in this case 12, which is known as 'top' (symbolised  $\top$ ), and a **unique least element**, here 1, known as 'bottom' (symbolised  $\perp$ ).

To grasp the second property necessary for latticehood, compare the following diagrams:

(4)



In (4a), a lattice, the set of elements greater than both  $c$  and  $d$  consists of  $b$  and  $\top$  ( $a$  is greater than  $c$  but not  $d$ ). Of this so-called 'up-set',  $b$  is the smallest element, and is known as the **least upper bound** or **join** of  $c$  and  $d$ , symbolised as  $c \vee d$ . Note crucially that  $c \vee d$  is unique. In (4b) on the other hand,  $c \vee d$  is not unique, and thus this is not a lattice. Here, the up-set of  $\{c, d\}$  is  $\{a, b, \top\}$ , where both  $a$  and  $b$  have 'equal claims', as it were, to being the least dominating element. In a lattice, then, every pair of elements must have a least upper bound. And this is not all — everything we have defined in terms of up-sets and the upper bounds can be defined dually for down-sets and lower bounds. The **greatest lower bound** or **meet** of  $\{a, b\}$  is symbolised  $a \wedge b$ , and is unique in (4a), our lattice, and non-unique in (4b), by similar arguments. To qualify as a lattice, then, all pairs of elements must have unique joins and meets.

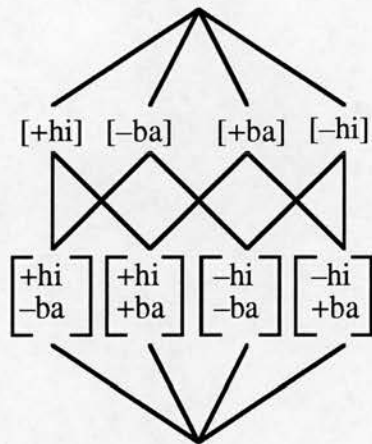
The diagrams we have employed above are known as 'Hasse diagrams', and it is as well to insist that such a diagram is not itself a lattice, but a *picture of a*

lattice. The *drawing* of lattices is as much an art as a science, as soon as the lattice attains a significant level of complexity. As Coleman and Local (1989) have insisted, considerable confusion results from the failure to distinguish phonological representations from *pictures* or *diagrams* of phonological representations. We enthusiastically endorse this view, and will strive to avoid contributing to such confusion in this thesis.

- *Stage 1: Lattice of possibilities*

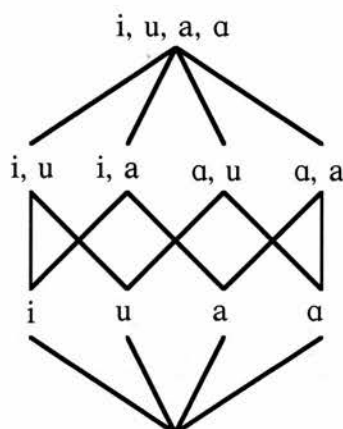
We can now begin to apply these ideas to the domain of phonology. As mentioned above, the set of descriptions in (1) form a partial order, which we diagram below:

(5)



Inspection of this diagram reveals a lattice structure, with partial descriptions — which have a more general denotation — positioned towards the top, and complete descriptions — which denote single segments — positioned towards the bottom. This lattice is derived directly from the lattice of denotations, ordered by set inclusion, given in (6), by simply relabelling the nodes:

(6)



$\perp$  in this lattice is the empty set, while  $T$  is the set of all segments in the inventory.

- *Stage 2: Lattice of representations*

This then is our starting point, a lattice containing all logically possible combinations of properties. It so happens that in Quileute every possible combination is attested, but consider now a more common and interesting situation. Suppose there were no front low vowel in Quileute. Two features would then be *more than* sufficient to describe the resulting three vowel system: specifically, the combination  $[-\text{high}, -\text{back}]$  is not attested and needs to be ruled out. Furthermore, if a segment is  $[-\text{high}]$  it is predictable that it will be  $[+\text{back}]$ , and thus redundancy is introduced. The new system (call it Quileute') is given below:

	i	ɑ	u
high	+	-	+
back	-	+	+

We now show how we express these facts in a lattice-theoretic approach. We begin as before, listing the denotations of all possible combinations of features:

(7)

$$[+hi] = \{i, u\}$$

$$[-hi] = \{a\}$$

$$[+ba] = \{u, a\}$$

$$[-ba] = \{i\}$$

$$\begin{bmatrix} +hi \\ +ba \end{bmatrix} = \{u\}$$

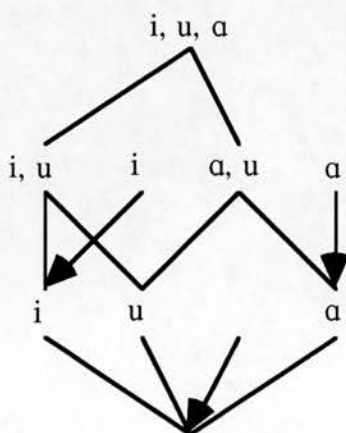
$$\begin{bmatrix} +hi \\ -ba \end{bmatrix} = \{i\}$$

$$\begin{bmatrix} -hi \\ +ba \end{bmatrix} = \{a\}$$

$$\begin{bmatrix} -hi \\ -ba \end{bmatrix} = \{ \}$$

Comparison with the corresponding list for Quileute shows that all we have done here is to remove the element /a/ from the denotations. This has the effect of altering the inclusion relationships. Notice first that some of the descriptions have become 'synonymous': different descriptions denote the same class of objects. Both [-high] and [-high, +back], for example, denote /a/. Indeed, this is a function of the fact that [+back] is redundant with respect to [-high]. Secondly, the specification [-high, -back] denotes the empty set, another way of saying it has no denotation: the description contains information which is incompatible (false) within this system. We can illustrate the way the inclusion relationships change with reference to the following diagram:

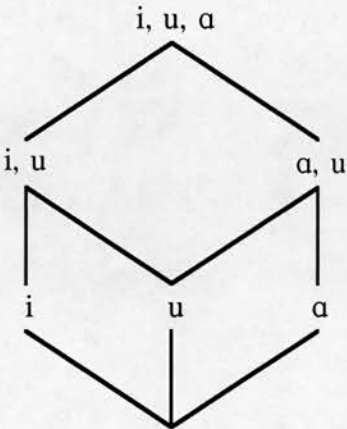
(8)





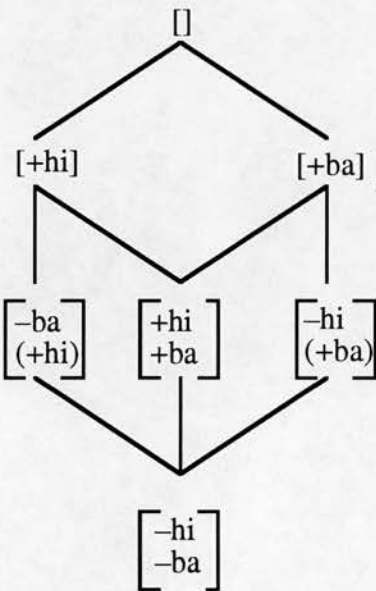
In effect, the lattice ‘collapses’, with all the nodes which once contained /a/ slipping down a rung: {a} to  $\perp$ , {i, a} to {i}, {a, a} to {a}. Stripping out the dead wood yields the following lattice:

(9)



And looking at this in terms of attributes:

(10)



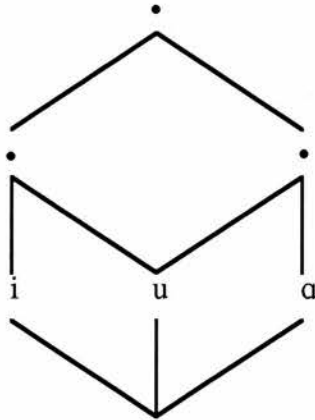
The label  $[-ba, (+hi)]$  simply abbreviates the pair of representations  $[-ba], [-ba, +hi]$ .

- *Stage 3: Lattice of features*

The third part of our exposition is concerned with the labelling of these lattices. The labels of the lattices in (9) and (10) are prolix in the same way as

is a diagram which includes all lines of dominance rather than just those expressing immediate dominance. For example, compare (9) above with the following:

(11)

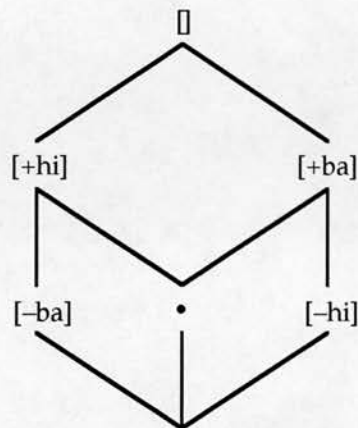


The labels for the unlabelled nodes can be derived by exploiting the structural relations provided by the lattice itself: the label of a node dominating labelled nodes is just the union of those labels; and given the transitivity of the ordering relation, this ensures that we can derive the label of any node, provided the lattice is appropriately ‘grounded’, with labels at the lowest level. In lattice theory, the nodes which cover  $\perp$  are known as **atoms** or **points** — Birkhoff (1948:7) cites Euclid ‘A point is that which has no parts’ — thus all we are claiming is that node labels be derived from the labels of their parts, removing superfluous labels as we removed superfluous lines. The significance of this move is revealed as we now move from denotations to descriptions.

The development of lattice theory followed from the publication in 1847 of Boole’s *Mathematical Analysis of Logic* (Birkhoff, 1948:iii). According to Boole, all reasoning can be reduced to the discussion of objects and their attributes. If attributes are combined using the logical operators ‘and’, ‘or’ and ‘not’, then the result forms a Boolean algebra, which has a lattice structure. Now it is natural to identify each attribute with the class of all objects having that attribute, and Boole’s second law stated that the algebra of attributes is a DUAL ISOMORPHISM of the algebra of classes. All this is by way of preparing the claim that whatever holds for the lattice of objects holds dually in the lattice of attributes. Thus we can omit attribute labels from the lattice without loss of information just as we omitted object labels. Just as, in the lattice of

objects, each node inherits the labels of the nodes it dominates, so in the lattice of attributes, each node inherits the labels of the nodes that dominate *it*. Compare (10) above with the following lattice, in which superfluous attributes have been omitted:

(12)



This simplification of lattice labelling is termed a REDUCTION by Wille and his associates in the Forschungsgruppe Begriffsanalyse (Ganter et al 1986, Luksch et al 1986).<sup>3</sup> In their terminology, each node in the lattice is associated with an EXTENT (in terms of its objects) and an INTENT (in terms of its attributes). The resulting minimally labelled lattice is also familiar in the field of knowledge representation, where it is known as an INHERITANCE HIERARCHY. We will develop this parallel in detail in subsequent chapters.

Now consider how this lattice expresses the redundancy in the system. In an account couched in terms of redundancy rules, this would be expressed by the following pair:

$$\begin{aligned} [-\text{back}] &\rightarrow [+ \text{high}] \\ [-\text{high}] &\rightarrow [+ \text{back}] \end{aligned}$$

And in an approach like Stanley's, couched in terms of conditions, the arrow would be interpreted as logical implication. In the lattice, on the other hand, the fact that [+high] is redundant with respect to [-back] is expressed by the dominance relation. The lattice thus yields an appropriate classification of objects (segments) and simultaneously expresses the implications between

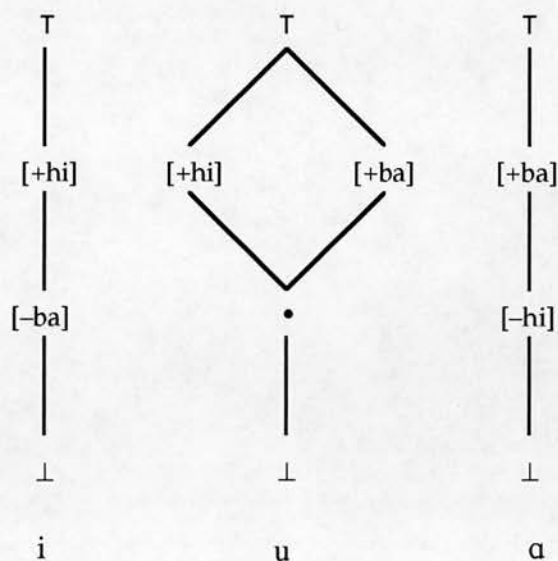
---

<sup>3</sup> For a rigorous formulation of the labelling strategy, see Davey and Priestley (1990:228).

attributes (features). Note also that each feature specification occurs just once in the lattice (resulting in four labels). This will always be the case: the lattice minimizes the incidence of feature specifications and maximizes the structural relations which hold between them. Such a lattice thus takes the place of an inventory matrix and associated redundancy statements, and is assured of expressing redundancy in the most efficient possible way, without superfluous repetition.

Note most importantly that redundancy is directly related to generality under this approach. Redundant features — predictable features — are located towards the top of the hierarchy; distinctive features — or ‘predictor’ features — are located at the bottom of the hierarchy. As our exposition has made clear, this lattice of redundancy is isomorphic to a lattice of classes ordered by inclusion, where top is more general and bottom is more specific. As the hierarchy reveals, then, the *distinctive* features are the *most specific* features. This result is extremely significant from the point of view of phonological theory, since it places specification within the general class of phenomena that can be treated in terms of the Elsewhere Condition. The implications of this will be examined in subsequent chapters.

If this is an inventory, what then is a segment? What kind of representation for individual segments do we derive from such an approach? As we will see subsequently — when we examine the structure of the lexicon that follows from this approach — there is in fact no need to consider representations apart from this lattice structure. But for the purposes of exposition, we can consider the representation of each of the three segments in Quileute’ to be as follows:



In lattice-theoretic terms, these diagrams represent the up-set associated with each atom of the lattice.

As promised, we have provided a representation which includes redundant specifications, but which encodes such specifications *as* redundant. It is immediately apparent from the form of the representation which features are distinctive and which are predictable. Again as promised, such representations concretely embody Stanley's insight: 'To say that a certain fully specified matrix is [...] redundant is actually to say that [...] its feature values are interrelated in ways determined by the constraints of the language'.

### 5.3. Partial Specification

The theory of structured specification resolves the problem of sensitizing phonological rules to the redundant/distinctive status of features without unwanted consequences. The most important aspect of the new formalism from this point of view, is that it distinguishes 'underspecification' from 'partial specification': the former is banished from the theory, replaced by structural relationships between features; the latter is permitted, and corresponds in a number of ways to the traditional notion of the archiphoneme. The correspondence is not complete, however, since this same confusion appears in a slightly different guise in the archiphonemic literature itself.



In the unification based framework, which makes extensive use of partial specification, it is this notion of 'archispecification' that is encoded. Work in unification-grammar has paid little attention to the role of redundant information, being almost exclusively concerned with syntax<sup>4</sup>. In phonology, however, lack of specifications has been used extensively to encode the notion of redundant specification. However, the decision to use blanks to code redundancy preempts the possibility of using blanks to encode archisegmental phenomena without the risk of ambiguity. We will examine the problems this preemption raises for current theoretical approaches to specification in Chapter 9.

We take it to be an advantage of the lattice-based approach to specification theory that we are able to offer a unified, formally explicit treatment of these two basic phonological notions which also expresses the fundamental difference between them. We will now explore these issues further, and examine how the kind of specification provided by the redundancy hierarchy relates to previous theories of specification.

#### 5.3.1. *Halle's formalization of the archiphoneme*

We have mentioned that Stanley (1967) objected to the use of blanks to code redundant specification. He also explicitly criticized the notion of the archiphoneme. His argument, however, is not directed against the notion itself, but against Halle's (1959) formalization of the idea. Interestingly, Halle did *not* attempt to express the notion of the archiphoneme by means of underspecification, even though underspecification was a central part of his theory. Rather, he attempted to exploit the structure provided by branching diagrams to express this idea. Thus even as early as Halle (1959) the formal expression of redundant features on the one hand, and of archiphonemes on the other, was in some measure distinct. As we now show, the specification hierarchy is able to express the notion of the archiphoneme in terms strikingly similar to Halle's attempt, but is immune to the criticisms voiced by Stanley.

Halle (1959:37) writes:

---

<sup>4</sup>The same is not true of unification-based lexicons, however, which have paid a great deal of attention to the treatment of redundancy. We discuss the parallels between this area of unification theory and our approach below.

The phonological system of a language will be presented by means of a branching diagram. Paths through the branching diagram starting at the initial node and terminating in one of its end points define different fully specified morphonemes. It will be shown later that segment types defined by paths starting at the initial node and ending at intermediate nodes — i.e., segment types that are “not different” from [non-distinct from MB] several fully specified morphonemes — play an important role in the functioning of language. These segment-types shall be called incompletely specified morphonemes...

And later (p.39):

It will have been noticed that incompletely specified morphonemes are analogous to the Prague school's ‘archi-phonemes’

The above formulation must be read with considerable care. The most important point to note is that ‘fully-specified morphonemes’ are *not* fully specified in the familiar sense: the interpretation is ‘fully-specified — for a morphoneme’.

Recall the details of Halle's model. He begins with MINIMALLY SPECIFIED representations: representations from which all predictable information has been omitted, in accordance with Condition (5). As we have seen, however, omitting all redundant information leads to representations which are non-distinct. Accordingly, certain MS rules must apply to fill in a certain amount of redundant information, in order to satisfy the distinctness condition. And as we have also seen, the choice of which information to include in order to achieve distinctness in the most economical way is overseen by the branching-diagram.

The branching-diagram thus monitors a process which is designed to render representations distinct; and at the termination of that process (i.e. when we have reached the end of the branches) then and only then will representations have achieved distinctive specification. They will not be any means *fully* specified — they will still contain numerous blanks which will subsequently be encountered by the phonological rules proper — they will be just sufficiently specified to be distinct. This is the stage of representation that Halle refers to as ‘fully-specified morphoneme’ — a more perspicuous description might be ‘fully-distinguished morphoneme’. If this ‘process of distinguishing’ is not carried out to the full, on the other hand — if we

terminate the process at some intermediate node — we will have a ‘partially-distinguished’ morphoneme: an archiphoneme.

The problem with this model is clearly that the two separate principles of specification are competing with each other — in fact, contradicting each other. On the one hand, lack of specification (Halle’s minimal specification) is designed to express the fact that the missing features are redundant — predictable from the features which are present. But simultaneously, lack of specification (Halle’s incomplete specification) is designed to express the fact that the missing features are distinctive.

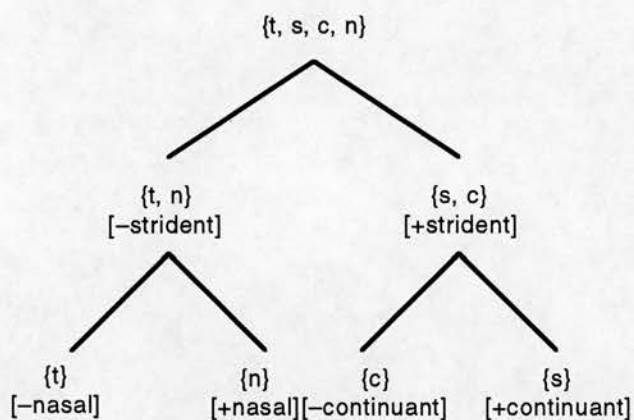
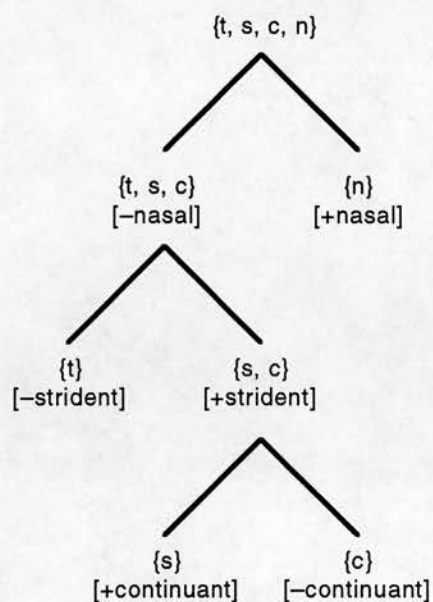
5.3.2. *Stanley’s criticism of the archiphoneme*

Stanley rejects Halle’s formalization of the archiphoneme on the following grounds:

the third reason for having branching diagrams [...] was that it was assumed that they formalize the notion of archi-phoneme. This assumption, however, is false. [...] it is easy to see that the branching diagram wrongly limits what the possible archiphonemes are [...] Thus branching diagrams fail to capture the notion of archiphoneme because of a quite basic and unavoidable fact of their structure

Stanley’s objection can be exemplified by comparing Halle’s two different branching diagrams for the example matrix cited in the previous chapter:]

	t	s	c	n
strident	–	+	+	–
nasal	–	–	–	+
continuant	–	+	–	–



The first diagram yields the following archiphonemes, corresponding to each non-terminal node of the hierarchy:

$A1 = \{t, s, c, n\}$

$A2 = \{t, s, c\}$

$A3 = \{s, c\}$

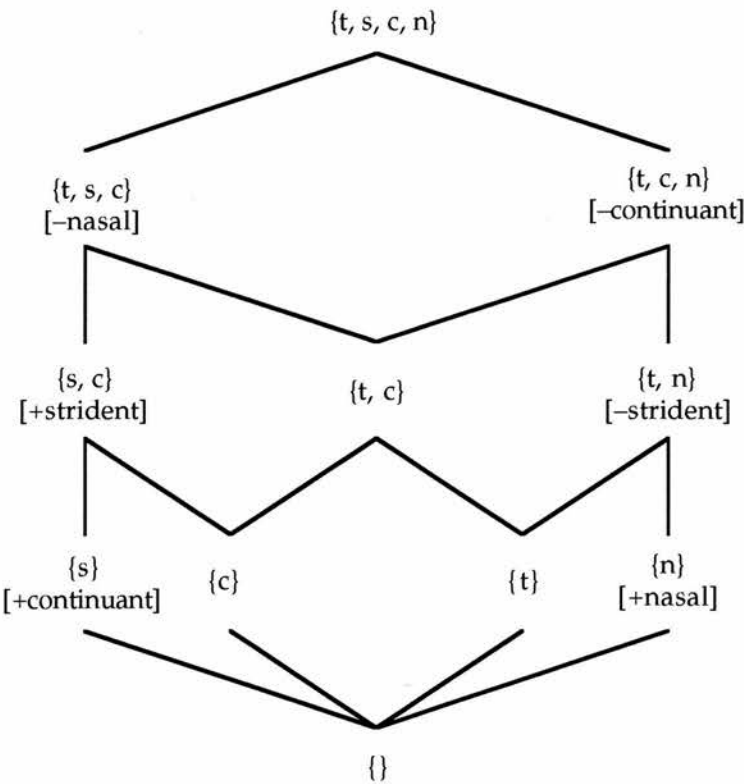
The second diagram yields the following archiphonemes:

- B1 = {t, s, c, n}
- B2 = {t, n}
- B3 = {s, c}

We might characterize the archiphoneme A2 as ‘oral’ ([–nasal]), and this clearly has as much right to archiphonemic status as any other element; yet this archiphoneme is not a possibility in the second branching diagram, the one claimed by Halle (1959:36) to provide the most efficient encoding of distinctiveness. Further, the ‘stop’ archiphoneme ([–cont]) is not an archiphoneme in *either* formulation. It is in this sense that the branching diagram incorrectly limits the archiphonemic possibilities.

### 5.3.3. Lattice-theory and the archiphoneme

We now reexamine the lattice-theoretic representation of this matrix from this point of view. We note first that there is a striking similarity in the notion of an archiphoneme defined by a ‘path starting at the root node and ending at an intermediate node’ in a hierarchy of features. In our approach, each node in the lattice can be interpreted as defining an archiphoneme. Here is the lattice we derive from Halle’s example inventory:





It is immediately apparent that the possibilities are not arbitrarily limited as they are in branching diagrams: the possibilities for archiphonemic status are in fact equivalent to the classes made available given the implicational constraints of the language.

#### 5.3.4. *Stanley's criticisms of branching diagrams*

As indicated in the quotation cited in the previous section, Stanley (1967) in fact identifies three main reasons why 'branching diagrams have been regarded as important' in generative phonology, and criticizes each in turn. We will discuss Stanley's objections in some detail, because they are, most conveniently, the single locus of argumentation responsible for the demise of branching diagrams in generative phonology (so persuasive were they); and also because, as we have argued, the redundancy hierarchy which we propose in this thesis bears such a striking resemblance to branching diagrams in a number of ways. We need to assure ourselves, then, that this new hierarchy is quite immune to the objections raised by Stanley.

One such function of branching diagrams — the formalization of the archiphoneme — we have just examined. The second function is, as we have discussed, that of ensuring that segments are rendered distinct. As Stanley observes, however, there are other ways of ensuring that segments are distinguishable besides subjecting them to Halle's procedure. As far as our approach is concerned, we simply note that the problem of non-distinctness arises from the decision to omit redundant information. We do not omit it, and thus do not face the problem.

It is the third and final function that Stanley (1967:408) gives most space to:

The branching diagram gives a hierarchy of features which can be interpreted as meaning that the features at high nodes (such as Consonantal) are in some sense more basic than the features at low nodes (such as Voiced)...

[This point] is more important, since there is obviously some kind of hierarchical relationship among the features which must somehow be captured in the theory.

The phrasing is interesting: 'some sense', 'some kind of hierarchy', 'must somehow be captured': for both Halle and Stanley, there appears to be an intuition here which it is not quite explicit. And Stanley immediately

questions whether the hierarchy provided by the branching diagram is in fact adequate to this intuitive notion:

an attempt to capture this hierarchy in a branching diagram seems somewhat strange in light of the fact that there may be considerable freedom in the way this branching diagram is constructed for a given set of systematic phonemes; a different choice of redundant feature values in this set will lead to a different branching diagram and thus to a different hierarchy of features

This then is the problem of INDETERMINACY which we discussed in the previous chapter. Stanley also argues that, if we simply adjust the branching diagram to yield the most intuitively satisfying hierarchy, we thereby admit that the structure is not independently motivated — does not follow from or correlate with any existing component of the phonology:

Obviously we would choose, if possible, that branching diagram which gives the hierarchy we feel is right; but this just means that we know beforehand what hierarchy we want and are simply choosing to represent it in a branching diagram. In this case, though, we might just as well describe the hierarchy separately since it has nothing essential to do with the branching diagram

In other words, if we strip the branching diagram of the practical function of rendering underlying representations distinct, then the vague sense that the diagram gives expression some intuitive hierarchy of features is insufficient reason to retain it as an essential part of the framework. Chomsky and Halle (1968) concurred, and branching diagrams disappeared from sight.

There are two issues here : (i) we require a hierarchy which is both determinate and independently motivated (ii) we want such a hierarchy to have a specific role to play in the theory of phonology, to 'buy us' something. As Stanley (1967:408) puts it:

What we want, of course, is to find some characteristics of features and their interrelations that force us, on independent grounds, to assume a PARTICULAR hierarchy of features. That is, we want to find some definite formal property of features, perhaps stated in terms of the different ways in which different features behave in the P rules or in the MS rules, which points in an unambiguous way to the existence of a specific hierarchy. Only in this way can a feature hierarchy come to represent more than a vague set of intuitions about what features are more basic than others, and thus only in this way can we discover a real

NECESSITY for incorporating the notion of feature hierarchy in the grammar

In the context of the model proposed in this thesis, Stanley's (1967:409) next observation is surprising:

What such a formal property might be is an open question at this point. We have seen, however, that the relation of redundant to non-redundant feature specifications in segments, which relationships can sometimes be stated in a branching diagram, is a formal property of features that is NOT, in itself, well enough determined on independent grounds to justify its use in characterizing a hierarchy.

However, as we have demonstrated in the first part of this chapter, the relation of redundant to non-redundant feature specifications *is* a well-determined, independent criterion — a characteristic of features and their interrelations — of precisely the kind that forces us to assume a particular hierarchy of features. Moreover, as subsequent empirical research has shown (which we discuss in a subsequent chapter) the formal property of features which this hierarchy defines — redundancy versus distinctiveness — gives rise to just that differential behaviour of features in the phonological rules and MS rules that Stanley conjectures (Steriade 1987, Clements 1987). Thus the lattice of redundancy, the 'specification hierarchy', independently and unambiguously defined in terms of the relationships between redundant and non-redundant features, is of precisely such a nature as to require incorporation into the theory of phonology.<sup>5</sup>

In conclusion, we see that none of Stanley's criticisms of branching diagrams carry over to the redundancy hierarchy as we have defined it. On the contrary, we find that this hierarchy has precisely those properties which Stanley would argue make it a *necessary* incorporation into the theory of phonology: it is built on characteristics of features and their interrelationships that force us, on independent grounds, to assume a *particular* hierarchy of features. And this characteristic — the predictable vs unpredictable nature of

---

<sup>5</sup> Recall that the branching diagram consists of features, some of which are unpredictable, others of which are predictable, all of which are distinctive. The hierarchical relationships defined over this mixed subset of feature specifications are not, as Stanley rightly states, determinate, nor can they be independently motivated. The redundancy hierarchy that we propose, on the other hand, consists of *all* features, and the hierarchical relationships defined over these features are the determinate and independently motivated relationships of logical implication.

feature specifications — does give rise to different behaviour with respect to phonological rules proper, a fact which is formalized in current phonological theory by the underspecification of predictable features. As we will see in subsequent chapters, there are numerous problems with such a formalization, problems which can be avoided by sensitizing phonological rules directly to these hierarchical relationships.

Further, we note that the problems Stanley raised with respect to branching diagrams follow from the decision first to omit redundant information, then to reintroduce a limited amount, then to define hierarchical relationships over just these features. If, on the contrary, we define hierarchical relationships directly over redundant and nonredundant features initially, the problems are avoided, and a determinate hierarchy emerges. We take this as corroboration of the fact that redundancy free representations play no role in the theory of phonology.

A final point before we conclude this discussion. The redundancy hierarchy also expresses the intuition that certain features are 'more basic' or 'more general' than others in a straightforward way. This is simply the fundamental notion of subclassification. Recall that the redundancy hierarchy expresses the logical implications between attributes simultaneously as it provides an appropriate classification of objects: a classification that combines subclassification and cross-classification in a lattice structure, sometimes referred to as a 'tangled' hierarchy. Thus the redundancy hierarchy may be interpreted dually as a CLASSIFICATION HIERARCHY. Inspection of the example lattice above reveals, for example, that the feature [+strident] is relevant to the classification of the subset {t, s, c}, and [+continuant] is relevant to the classification of the sub-subset {s, c}. Part of the problem experienced by Halle in determining such a classification hierarchy is that the appropriate primitive is not a *feature* (such as [continuant]) but a *feature specification* (such as [+continuant]).<sup>6</sup> Another problem was the arbitrary limitation to a tree-structured hierarchy, appropriate for expressing *only* subclassification, and thus obscuring the actual classificatory relationships.

---

<sup>6</sup> As we will see in the next chapter, this is related to the misapprehension that the system of categories made available to the phonology is Boolean.



The example above is not extensive enough to display an intuitively convincing notion of generality, but as will become clear subsequently when a larger number of segments and features are incorporated, features like [consonantal] will indeed gravitate toward the top of the lattice (being relevant to the classification of a large number of segments) while features like [distributed] will fall to the bottom (being relevant to the classification of only a relatively small number of segments). This notion of subclassification went unexpressed in the framework of SPE, where all classification is cross-classification. More recently, an attempt has been made to rectify this failing within the framework of FEATURE GEOMETRY. In a subsequent chapter, we will argue that the attempt to press the categories of feature geometry into such service is misguided, and that it is in fact the redundancy hierarchy that provides the appropriate locus for such generalizations.

#### 5.4. Default Specification

We have examined in considerable detail the framework of STRUCTURED SPECIFICATION, in which redundant values are present in underlying representations, with their redundant status coded by the structural relationships they bear to other co-occurring features. We have also shown how the same framework formally expresses the notion of PARTIAL SPECIFICATION, and how non-terminal nodes of the lattice correspond in significant ways to the traditional notion of the archiphoneme. We close this chapter with an introduction to the notion of DEFAULT SPECIFICATION.

Consider the full specifications for the vowels of Okpe:

	i	ɪ	e	ɛ	a	ɔ	o	u	ʊ
back	–	–	–	–	+	+	+	+	+
round	–	–	–	–	–	+	+	+	+
high	+	+	–	–	–	–	–	+	+
low	–	–	–	–	+	–	–	–	–
ATR	+	–	+	–	–	–	+	–	+

In radical underspecification theory, the default or unmarked value for each feature is literally unmarked, omitted from the representations which form the inputs to phonological rules. One reasonable set of default settings would be given by the following set of redundancy rules:

- [] → [–round]
- [] → [–high]



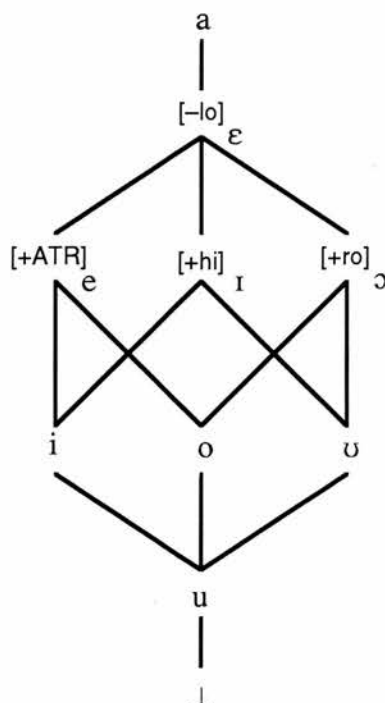
[] → [+low]

[] → [-ATR]

which give rise to the following underspecified representations:

	i	ɪ	e	ɛ	a	ɔ	o	ʊ	u
round						+	+	+	+
high	+	+						+	+
low	-	-	-	-		-	-	-	-
ATR	+		+				+		+

A cursory inspection of the table above reveals an information ordering, from the least specified segment /a/, to the most specified segment /u/. The ordering is not total, however: /ɪ/ and /e/, for example, while both more specified than /ɛ/, are not more specified one than the other. Here, degree of specification corresponds to degree of markedness. /a/ is exceptionless, exhibiting the default value for every feature; /u/, on the other hand, has the marked or exceptional value for every feature. The partial ordering that motivates the omitted specifications can be represented directly, as in the following lattice:



The relative ordering of /ɛ/, /e/ and /i/ are clearly expressed in the lattice, as are the complete set of such orderings. Intuitively, this lattice represents

the ordering of the elements in terms of exceptionality: the least marked or default vowel is at the top, the most marked (most exceptional) vowels are at the bottom. Just as in the lattice of redundancy, this lattice represents an inheritance hierarchy: the node labelled /i/, for example, inherits both the features [+ATR] and [+hi]; the node labelled /u/ inherits all marked features.

The lattice above shows only *marked* features. Unmarked features can be incorporated into the lattice by drawing on ideas developed in the study of INHERITANCE SYSTEMS in knowledge representation theories, a program which we will pursue in greater detail in a subsequent chapter. For the moment, we simply wish underline this key point: just as redundancy exhibits a SUBSUMPTION STRUCTURE, with more general (predictable) properties subsuming more specific (unpredictable) properties, so too markedness exhibits a subsumption structure, with more general (default) properties subsuming more specific (exceptional) properties.

## 6. Aspects of structured specification

---

### 6.1. Introduction

In this chapter we explore some of the immediate implications of adopting a structured specification approach. In the process, we introduce terminology which will be useful in subsequent chapters, and in our solution to the problem of reciprocal dependencies in section 6.3.

### 6.2. Utilizing Lattice-theoretic ideas

A potential advantage of forging links between apparently disparate fields of research is that we can use results which have been established in one field to illuminate issues in the other. The lattices investigated by pure mathematicians typically have a more regular structure than that exhibited by the hierarchical relationships found in the empirical sciences. Nevertheless, in this section we take some first steps in this direction, introducing certain lattice-theoretic notions that seem to have direct application to phonological relationships.

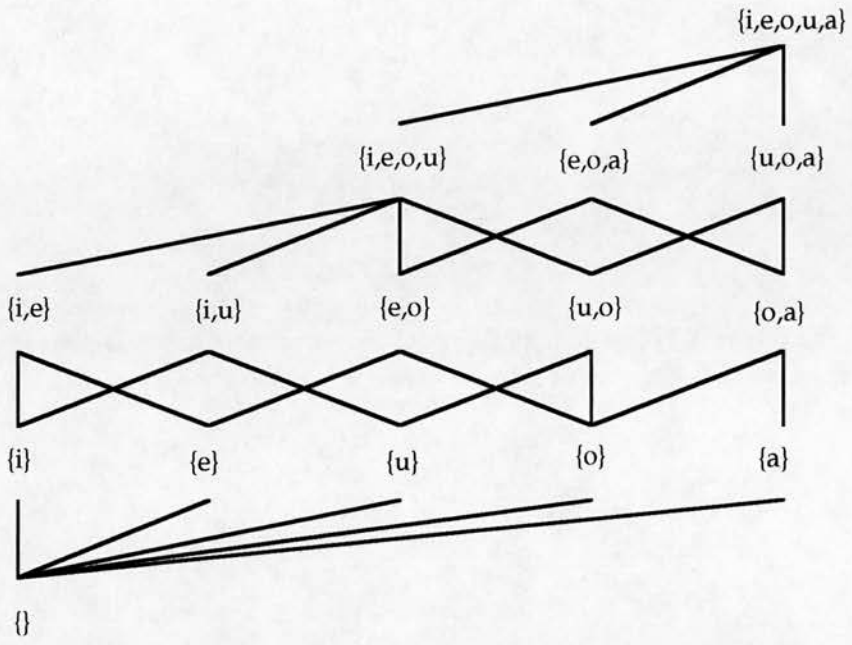
For the purposes of exemplification, let us consider a plausible example of moderate complexity: the canonical five-vowel inventory, as represented in the following table:

(1)

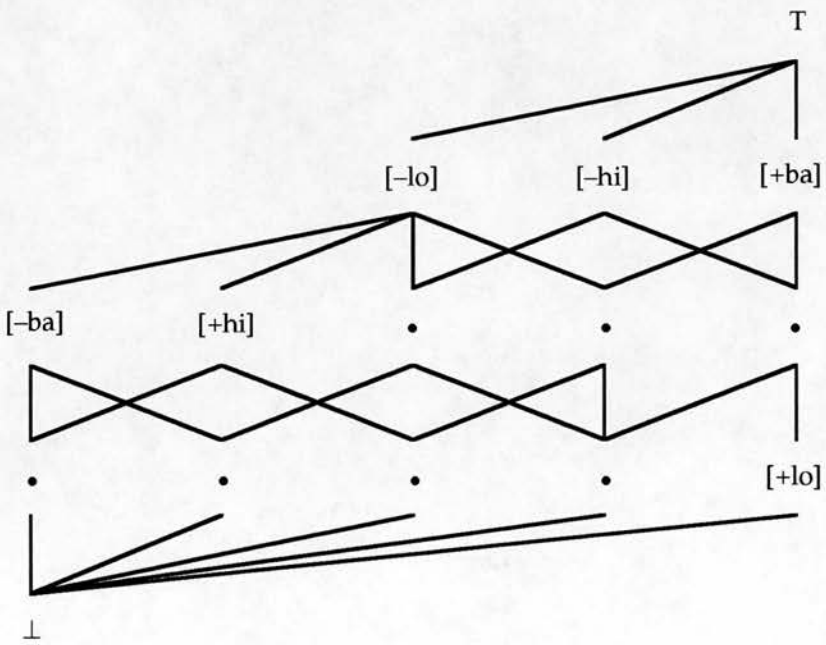
	i	e	a	o	u
high	+	–	–	–	+
low	–	–	+	–	–
back	–	–	+	+	–

A lattice can be created from this table employing the method described in the previous chapter. We first form the set of all possible feature matrices — both fully specified and partially specified — that can be created from these features. We then determine the denotation of each matrix, and arrange the denotations in a lattice based on the relation of set inclusion. The resulting lattice is given below, first in terms of natural classes, then in terms of distinctive features:

(2)



(3)



### 6.2.1. Boolean lattices and negation

We have mentioned that Boolean algebras were the first lattices to be investigated, and their mathematical properties are well-known. In view of the prevalence of so-called 'Boolean-valued' features in generative phonology, it is interesting to consider whether the lattices that result from the relations between phonological features are Boolean. It turns out they are not: a result which has implications both for phonological theory, and for the logical formalization of phonological theory. First, then, we examine the definition of a Boolean lattice.

As a preliminary step, we need to establish the notion of a distributive lattice. A distributive lattice is a lattice satisfying an extra condition:

**Definition:** A lattice is said to be DISTRIBUTIVE if it satisfies the distributive law:  $(\forall a, b, c \in L) a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$

Any lattice of sets (a lattice whose elements are closed under intersection and union) is distributive. Any powerset lattice  $\wp(X)$  is distributive.

A Boolean algebra is a distributive lattice with additional structure that mimics the complementation in a powerset. We now define the notion of complement in lattice theoretic terms:

**Definition:** Let  $L$  be a lattice with  $\perp$  and  $\top$ . For  $a \in L$ , we say  $b \in L$  is a COMPLEMENT of  $a$  if  $a \wedge b = \perp$  and  $a \vee b = \top$ . If  $a$  has a unique complement, we denote this complement by  $a'$ .

In a distributive lattice (and hence in a Boolean lattice) an element can have at most one complement. A lattice element *may* have no complement. In a lattice  $L$  of subsets of a set  $X$ , the element  $A$  has a complement iff  $X \setminus A$  belongs to  $L$ . We may now define a Boolean lattice:

**Definition:** A lattice  $L$  is called a BOOLEAN LATTICE if:

- (i)  $L$  is distributive
- (ii)  $L$  has  $\perp$  and  $\top$ ,
- (iii) each  $a \in L$  has a (necessarily unique) complement  $a' \in L$ .

It is apparent that the lattice diagrammed in (2, 3) is *not* Boolean. This can be established directly by inspection, noting that the lattice element  $N = \{e, o\}$  has



no complement in the lattice: the set  $N' = \{i, a, u\}$  does not occur. This is an important property of phonological lattices, and one of the empirical constraints that distinguishes the lattices of phonology from the lattices of logic. The reason that  $\{i, a, u\}$  is not part of the lattice, of course, is that this set is not a NATURAL CLASS. And in general, it is not necessarily the case that the complement of a natural class is itself a natural class. The unnatural set could of course be expressed if we allowed a negation operator to form part of phonological categories, such as:

$$\neg \begin{bmatrix} -\text{hi} \\ -\text{lo} \end{bmatrix}$$

but as we can see, this would introduce extra and unwanted power into the system.

Thus not all categories (elements) in the lattice have complements. Inspection of the lattice in (2, 3) also reveals, however, that categories consisting of single (binary) feature specifications do have complements: as we would expect, the lattice-theoretic complement of the node labelled [+high] is the node labelled [-high], and so on. To this extent, the features so labelled are seen to be Boolean. However, it seems odd to adopt a system for which the negative of a single feature, whatever it is, necessarily constitutes a natural class, whereas the negative of any conjunction of features does not. This is an empirical, not a logical matter. By the law of the excluded middle, it is logically necessary that any object which does not have a property lacks it: it is not phonologically necessary, however, that this set should be referable by phonological rules. It is just as much a logical necessity that a set which does not have a *conjunction* of properties lacks that conjunction, but as we have seen, there is no presumption that such a class forms part of phonological theory.

And, in principle, there is no reason to assume that the set of objects which are [-F] form a natural class just because the set of objects which are [+F] do. Indeed, as we will see in the next chapter, one of the claims of the theory of feature geometry is that it is able to rectify precisely this problem inherent in the feature system of SPE, through the use of monovalent articulator nodes. It must be stressed, however, that this is not a failure of the notion of classification itself. We do not have to take an overly concrete, 'non-classificatory' view of features in order to avoid the problem. Given that

phonological *systems* are not Boolean, for the reasons we have described, there is no reason to assume that *feature specification* is necessarily Boolean. Indeed, we should rather assume it is not. There may indeed be instances where both the presence of some feature, and the absence of that feature, constitutes a natural class, but this should be viewed as a contingent, rather than a necessary fact, of phonological systems.

If indeed such a relation holds between features, it will be reflected in the lattice structure of relations holding between them, and need not be represented in the syntax of the feature 'name'. That is, the feature specifications [+high] and [-high] would be complements ('Boolean') in the above lattice even if they were named HIGH and LOW (or FOO and BAR). By explicitly encoding lattice structure (redundancy structure) in phonological representations, we obviate the need for two-part feature specifications.

This also provides a new perspective on the notion that a feature may be monovalent in one language but bivalent in another. This amounts to the question whether the extension of a feature is disjoint from the extension of some other feature in a particular inventory. In other words, 'Booleanness' is not an *inherent* property of a feature, but a function of the *relation* that feature bears to other features in the system: a feature will be 'Boolean' just in case it has a complement in the lattice of redundancy.

The position that follows most naturally results from these considerations is as follows. Given that the notion of natural class is predicated on the notion that one cannot refer to the absence of a conjunction-of-features, we assume that one cannot refer to the absence of a single feature. Just as in the case of negated conjunction, if we need to refer to such a class, we need either to find a property that positively identifies it, or state a disjunction. This has the effect of making all features 'unary'; replacing, say, [+continuant] and [-continuant] with the features FRICATIVE and STOP. However, this is *not* to say that such features cannot be Boolean: they may. Further, they may be Boolean in one language and not in another. The property 'Boolean vs non-Boolean' is in fact orthogonal to the notion of 'presence vs absence' of feature. Indeed, as we will see below, extension of lattice-theoretic notions raises the possibility that features may be 'pseudo-Boolean' in a precise mathematical sense.

Some confusion has recently arisen in the phonological literature regarding 'unary' and 'binary' (Boolean) classification. Thus we find claims such as the following (Paradis and Prunet 1991:23):

Reference to [0F] creates potential ternary distinctions only if [F] is a binary feature [...] However, class nodes, such as articulators or PNs [place nodes] are unary. Yip (1989:370) points out that with unary nodes, distinguishing a gap and a specification does not create the same problems as with binary features since only presence or absence of a given node is available

However, if it is undesirable that binary features be capable of making ternary distinctions, then surely it is undesirable that unary features be capable of making binary distinctions, and for precisely the same reasons. If this is not the case, in what sense are these features 'unary' at all? (As we will see in the next chapter, there is a further confusion here, regarding the categories place node and articulator node). Paradis and Prunet (1991:23) also claim that:

it is clear [...] that a PN gap must be distinguished from a PN specification [...] The possibility of referring to a gap was hinted at by Kiparsky (1985:98-99), but some current positions [...] do not consider that incomplete structures can be referred to by MSC's. Their argumentation is based indirectly on the notion of distinctness

If it is possible to 'refer to a gap', the absence of a node, and that node picks out a cluster of properties, then this is equivalent to negating feature-matrices, as has been prohibited in generative phonology — and for good reason, since it enables us to refer to unnatural classes. It might be objected that this strategy is 'made safe' by the extra machinery of feature-geometry, since the node whose absence is referred to dominates a 'natural-class of features'; thus we cannot negate *arbitrary* conjuncts of features. But as we have seen, the negation of a natural class is not necessarily a natural class.

In short, the varieties of specification which result from combining the hierarchy of feature geometry, unary and binary features, and underspecification, have raised questions about the proper interpretation of the resulting structures, and have reopened theoretical debates concerning the fundamental notions of distinctness and rule application; debates which are often undermined by a theoretical confusion regarding the fundamental properties of classification. The structured specification approach locates

discussions of classification and distinctness within the framework of mathematical logic and algebra, and is thereby able to illuminate these issues and forestall possible confusions. We return to this theme when we consider the implications of our approach for the theories of feature geometry and underspecification theory in subsequent chapters.

### 6.2.2. *Pseudo-complements and intuitionistic negation*

We have noted that the lattice of phonological categories is not Boolean, since not every element of the lattice has a complement. We have seen how the introduction of a negation operator would permit the construction of a complement to every class, but that such classes are not 'natural' in the phonological sense. The logic of phonology is, as it were, more constrained than a Boolean logic. Within lattice theory, however, a relation has been defined which is weaker than that of strict complementation, and which may have application to phonological description. Investigation of this issue will require empirical study of phonological inventories. Our aim here is simply introduce the relevant theoretical notion as a basis for further work.

The lattice-theoretic notion PSEUDOCOMPLEMENT is defined formally as follows:

**Definition:** For  $a, b \in L$ , we say  $a^*$  is the PSEUDOCOMPLEMENT of  $a$  if

$$a^* = \max \{b \in L \mid b \wedge a = \perp\}$$

In other words, a given node may have a number of elements with which it unifies to yield  $\perp$ : if there is a unique greatest element of this set, it constitutes the pseudocomplement. Thus the pseudocomplement of  $a$ , is just the greatest element disjoint from  $a$ , in case such an element exists. A lattice is said to be PSEUDOCOMPLEMENTED if all its elements have a pseudocomplement.

Just as strict complementation is a defining characteristic of Boolean logic, so the notion of pseudocomplementation is connected with an alternative but important logic. Brouwer and the intuitionist school rejected the claim that the disproof of 'not-p' implied the truth of 'p'. As Birkhoff puts it (1948:194) 'Such negative proofs seem particularly unsatisfactory when 'p' asserts the existence of a number, but the disproof of 'not-p' indicates no procedure for finding it'. Brouwer and his school rejected 'non-constructive' proofs. Motivated by such metaphysical considerations, an algebra of logic was



developed in which the biconditional  $x \leftrightarrow \neg(\neg x)$  is denied, and replaced by the weaker conditional  $x \rightarrow \neg(\neg x)$ . In lattice-theoretic terms, the equality  $(x')' = x$  is replaced by the weaker inequality  $(x')' \leq x$ . This then is Heyting's formulation of intuitionist or Brouwerian logic, which may be phrased as follows (Birkhoff 1948:147):

**Definition:** A BROUWERIAN LOGIC is the dual of a relatively pseudo-complemented lattice.

Birkhoff's definition (p. 147) of pseudocomplemented lattices is in fact the dual of the definition of a Brouwerian logic due to McKinsey and Tarski.

We mentioned above that the fact that segment inventories do not constitute a Boolean algebra has implications for the logical formalization of phonology. In many extensions of unification-based approaches to syntax, it has been suggested that unification be augmented with other operations or algebraic relations (Shieber 1986:63) such as disjunction and, significantly for our purposes, negation. Coleman (1991) explicitly employs a negation operator in his unification-based phonological formalism. There are a number of theoretical difficulties involved in such an extension<sup>1</sup>, but leaving these aside, our discussion above shows that such a move should not be taken lightly in the phonological domain, if we wish to preserve an essential insight of phonological theory, the natural class. Our discussion also suggests that, if it is found that negation (lattice-theoretic complementation) of some sort must be introduced into the formalism, a weaker version than that found in Boolean algebra might be considered: intuitionistic negation (pseudo-complementation). There is in fact a precedent for such an extension to the unification formalism: Pollard and Sag (1987:43-44) employ the pseudo-complement operator (and the associated Heyting algebra) to provide a formal definition of their CONDITIONAL feature structure. And in the logic employed by Moshier and Rounds (1987) for describing partially specified

---

<sup>1</sup>Current approaches to this problem do not admit disjunction in feature-structures at all, but crucially define a specialized description language for which the syntactic objects (feature-structures) supply the semantics, and use disjunction in the *descriptions* of feature-structures which the language makes available. See (Johnson 1988) for one such approach and an overview.



data structures — a logic intended precisely to incorporate negation into a unification-based framework — negation is interpreted intuitionistically.

The role of such an operator in phonology — and the extent to which the appropriate logic is intuitionistic — is a question for future research.

However, it is interesting to speculate that the objections that have been raised to negative specifications in phonology (by, for example, researchers in Government Phonology) are analogous to the intuitionistic objection to negative (non-constructive) proof. For example, such a position might reject the idea that recognition could proceed by identification of the absence of some phonological property<sup>2</sup>. Such a position could be characterized as ‘intuitionistic phonology’; and to the extent that, in a logical approach to phonology, ‘recognition is deduction’, is equivalent to requiring that recognition employ a constructive proof, and that the algebra of features constitute a Heyting algebra.

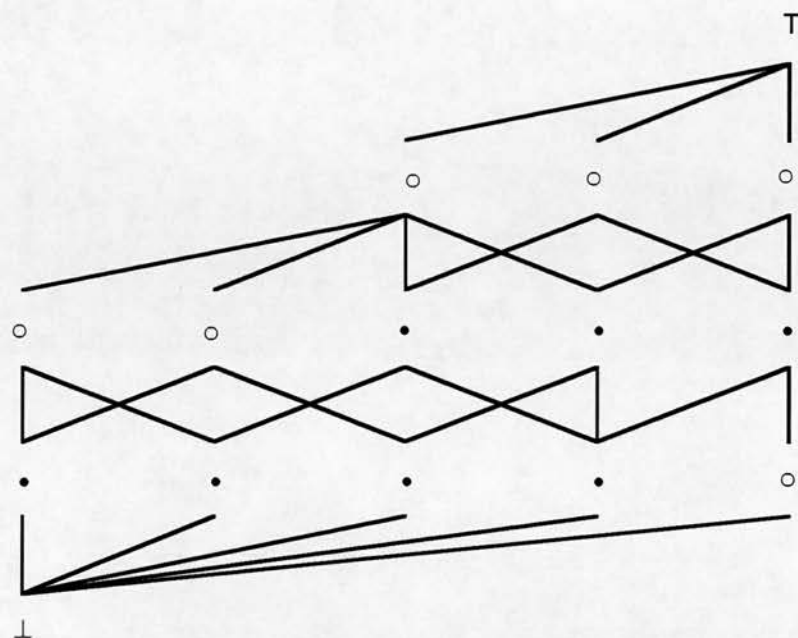
### 6.2.3. *Irreducible elements*

It will be noted that the distribution of feature specifications in our example lattice (2, 3) corresponds to a ‘geometrical’ property of the lattice:

---

<sup>2</sup>As Jonathon Kaye once graphically demonstrated during a discussion of ATR harmony in an Edinburgh restaurant, how do you spread [–ketchup]?

(4)



The nodes which are annotated with a feature (symbolized here by the open circles) are just those nodes which are dominated by a single line, or more technically, 'covered by a single element'. This is a well-known structural configuration in lattice theory: an element which has exactly one upper cover is known as MEET-IRREDUCIBLE (Davey and Priestley 1990:165-6). In other words, such a node cannot be expressed as — cannot be *reduced* to — the meet of two other nodes. Given that the basis of our ordering is set inclusion, the interpretation of this property is that the set located at this node cannot be expressed as the intersection of two other sets. It follows that the only way to define such a subset is to predicate another feature which distinguishes that subset.

Meet-irreducible elements (and their dual, join-irreducible elements) are extremely important in lattice theory. We have already introduced the notion of Boolean lattices, and the notion of the atoms of a lattice (the set of elements which cover  $\perp$ ). The archetypal example of a Boolean lattice is the power set  $\wp(X)$ : the set of classes of Quileute vowels was such a lattice. Every element of such a lattice is a union of singleton sets (atoms). Thus the Boolean algebra  $\wp(X)$  may be regarded as being 'built up from' the set of atoms of  $L$ . While

atoms are the appropriate building blocks for Boolean algebras, the representation of finite distributive lattices (a much richer class) requires a more general notion: the ‘atoms’ for such cases are provided by the irreducible elements. The set of such elements can be thought of as a kind of ‘skeleton’ or ‘representing set’ from which the lattice may be reconstructed. In the general case, the lattice can be reconstructed only by a rather complicated process from the union of meet-irreducible and join-irreducible elements. But, in the case of a (finite) *distributive* lattice, the lattice may be reconstructed from the meet-irreducible (or join-irreducible) elements alone. In fact, the lattice is equivalent to the family of all down-sets of join-irreducible elements (Davey and Priestley 1990:169). We note the satisfactory result that there is a congruence between the distribution of distinctive features and the representing set of the lattice, reinforcing the notion that the set of available combinations of properties — the set of phonological representations — is derivable from the implicational relationships which obtain between those properties.<sup>3</sup>

The notion of irreducible element in fact allows us to express a second, more complex type of redundancy relation in terms of the structural configurations of features in the lattice. This will be crucial in our treatment of reciprocal dependencies, and it is to this we now turn.

### 6.3. Reciprocal dependencies

We now show how our approach solves the problem of reciprocal dependencies in the segment-structure case. The treatment of sequence-structure dependencies will be addressed subsequently, but is a straightforward extension of this context-free case. Our solution is simple: since we omit no features, we are not forced to arbitrarily omit one or other term of a reciprocal dependency. We *express* the mutual aspect of the dependency by the special nature of the relation – the particular hierarchical organization – that obtains between such features.

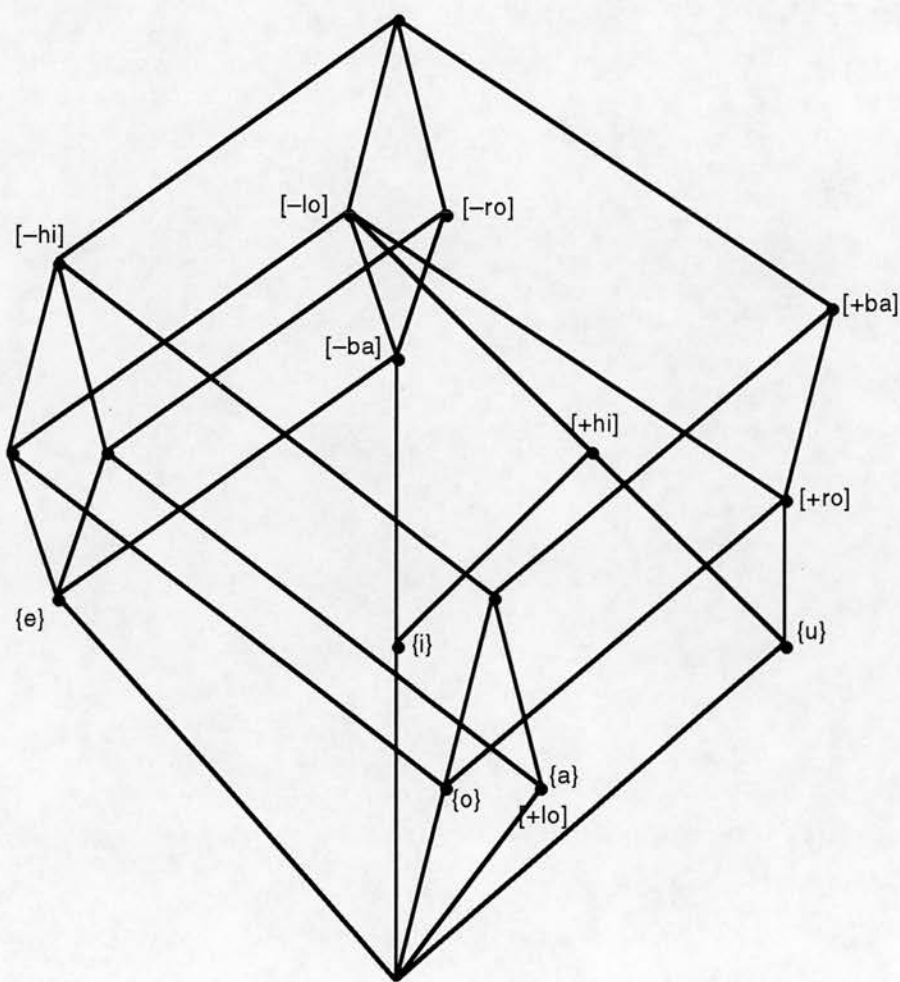
As mentioned in Chapter 4, It is well-known that in the canonical five vowel system either [back] or [round] is sufficient to distinguish all segments. This is

---

<sup>3</sup>The interested reader should refer to Davey and Priestley (1990:163–183) for proofs and elaboration of the claims made in this section.

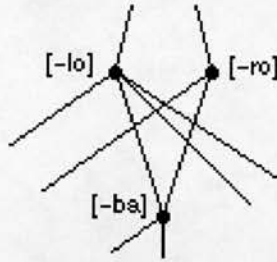
also apparent from the lattice in (3) above, which simply uses the feature [back]. There is thus a reciprocal dependency between these features, and the standard approach forces us to omit one or other feature from underlying representations. At various times, arguments have been advanced to omit one in favour of the other, motivated in terms of the way in which segments pattern in phonological rules, but even when no motivation can be offered, and the features co-vary in a parallel manner, the standard theory forces an arbitrary choice. In our approach, both features may be included in representations. The five vowels classified by the features [high], [low], [back] and [round], may be diagrammed as follows:

(5)



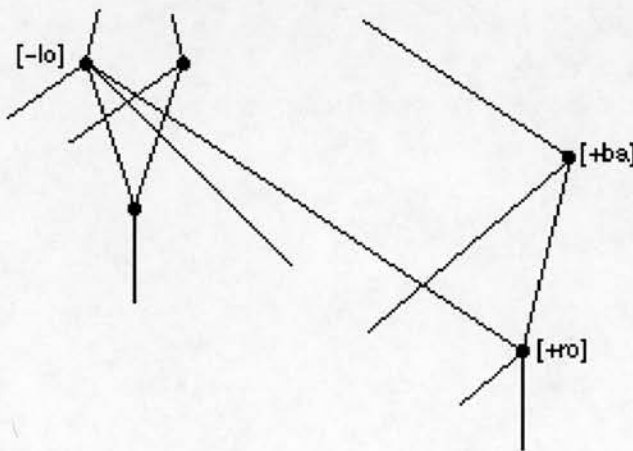
Consider the following fragment of the lattice:

(6)



The fact that [-round] is redundant with respect to [-back] is expressed in the lattice by the fact that [-round] dominates [-back], in just the same way that the redundancy of [-low] with respect to [-back] is expressed. This notion is familiar to us from the discussion in the previous chapter. But equally, the fact that the [-back] is itself redundant given the *conjunction* [-low, -round] is expressed by the fact that [-back] labels a *meet-reducible* element. To adopt a procedural metaphor, given the information [-low, -round], the *addition* of the information [-back] does not 'move us down' the network — does not provide any discriminatory power to identify a smaller class than the present one. In other words, the information is redundant. Exactly the same arguments hold of the fragment involving the other values of [back] and [round]:

(7)



Note the structural position of the node [-low] in each of these configurations, and recall the feature matrix associated with this set:



	i	e	o	u	a
low	–	–	–	–	+
round	–	–	+	+	–
back	–	–	+	+	+

The lattice structure reflects the fact that asymmetries in the /a/'s values for [back] and [round] leads to asymmetries in the implicational relationships. That is, the implication  $[+round] \supset [+back]$  holds throughout the inventory, 'context-free' as it were; whereas the second half of the biconditional  $[+back] \supset [+round]$  holds just in the case of  $[-low]$  vowels. That is, *given*  $[-low]$  ('in the context'  $[-low]$ ), *then*  $[+round]$  is implied by  $[+back]$ . Similar arguments hold in the case of  $[-back]$  and  $[-round]$ . Thus we see that the two structural types of redundancy coded by the lattice correspond to 'context-free' and 'context-sensitive' redundancy statements.

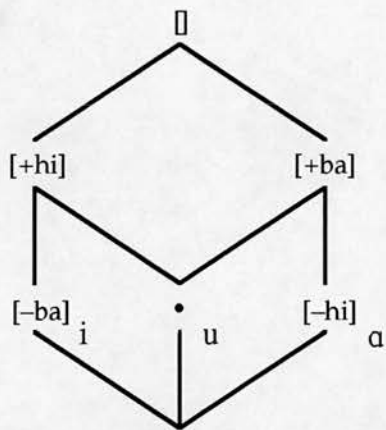
If we compare the lattice in (5) above which employs [back] and [round] with the lattice over the same inventory which employs only [back] ((3) above), we see that the result of adding an extra binary feature is not just that more nodes are created (corresponding to new classes that can be defined by reference to one value of the feature) but that also (i) meet-reducible elements are labelled, corresponding to features which define a class which is independently defined by a combination of existing properties, and (ii) irreducible elements become reducible ( $[-back]$  for example) corresponding to classes which can now be characterized by a combination of new and existing features. What is also revealed is an asymmetry between positive and negative specifications for certain features: thus while  $[+round]$  is reducible,  $[-round]$  is irreducible; and given  $[-round]$ ,  $[-back]$  is reducible, while  $[+back]$  remains irreducible. Note then, that even when redundant information of this nature is added to the lattice, the set of irreducible feature specifications is still well-defined, but the new and complex redundancy relationships which arise — even those involving reciprocal dependency — can be elegantly expressed in terms of structured specification. This is an important result, since as the number of features classifying a particular space increases, so will the instances of reciprocal dependency. Indeed, the phenomenon of reciprocal dependency is a pervasive and fundamental property of any system of redundancy. Given that the redundancy of phonological systems serves the important purpose of aiding recognition, then we might expect that an inherent structural property such as mutual predictability be exploited to the full. We return to this issue

when we consider the implications of our approach to speech recognition in a subsequent chapter.

#### 6.4. The hierarchy of features and gradient processes

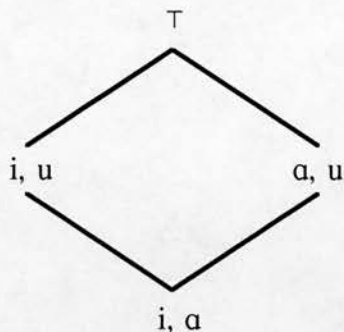
Yet another way to interpret the hierarchy we propose is in terms of SIMILARITY. In other words, the lattice is capable of answering such questions as: Which are more similar, /a/ and /o/, or /a/ and /i/? To see this, consider the limiting cases. First consider two segments which differ in every way, which have no features in common: the join of two such elements will be top. Now consider two elements which differ solely in terms of a single specification. The join of two such elements will cover (immediately dominate) each of them. Consider, for example, Quileute':

(8)



Intuitively, the segments /i/ ([+high, -back]) and /a/ ([-high, +back]) are totally dissimilar. The segments /i/ and /u/, on the other hand, are more similar, in that they are both [+high]; equally the segments /u/ and /a/ are similar, in that they are both [+back]. However, of the two similar pairs, neither is more similar than the other. Thus we have a partial order SIMILAR-TO defined over pairs, as expressed in the following diagram (most-similar (identical) is top and dissimilar is bottom; note this diagram (a lattice!) is not part of our theory, but simply illustrative):

(9)



However, this ordering can be derived directly from the redundancy hierarchy (8): the join of  $x$  and a *similar* element will be less-than the join of  $x$  and a *dis-similar* element. Thus the join of /i/ and /u/ is less than the join of /i/ and /a/. However, this order on joins is not always defined (it is a partial order): the joins of /u/ and /a/ on the one hand, and /u/ and /i/ on the other, are not ordered — the desired result.

Given that the representation explicitly encodes such a ranking, is there evidence that phonological rules are sensitive to it? An interesting case of precisely this type has recently been discussed by Pierrehumbert (1992) regarding Arabic dissimilation.

Her discussion is based on Lightner's (1973) observation that homorganic consonants which differ in many features combine more freely in Arabic than ones which differ in only a few features. That is, the strength of the effect is related to the overall similarity of the target segments. Arabic thus exhibits a *gradient OCP* effect: the OCP applies only to consonants which are perceived to be similar, and the strength of the effect increases with perceived similarity. Identical consonants are viewed as maximally similar, and so the total OCP arises as the limiting case. Pierrehumbert argues that Arabic differs from other languages in its relatively low similarity threshold for the effect to be active: even rather dissimilar combinations of homorganic consonants, such as /f/ and /m/, are found less often than might be expected. A precise comparison of the 'similarity gradient' uncovered by Pierrehumbert and the similarity ordering implicit in the lattice-theoretic approach is a matter for

future research; but in principle the two approaches exhibit a striking correlation.

Pierrehumbert rejects unreconstructed full-specification theory of the SPE type, since the principle that every segment is specified for a feature that is relevant to any segment fails to capture the relevant similarity ranking. She adopts a version of contrastive underspecification, and defines a similarity ranking in terms of the specified features. In the structured specification approach, however, the relevant hierarchy is already defined, and is seen to be structurally related to the independently motivated redundancy hierarchy.

Pierrehumbert's results are based on a statistical study of the Arabic lexicon, and clearly the gradient pattern revealed is predicated on this type of quantitative analysis — analysis which access to computerized corpora is making increasingly viable. Investigation of statistical correlates of the hierarchy we propose thus seems to us a promising field of future research.

# 7. Structured specification and feature-geometry

---

## 7.1. Subclassification and valency

One of the advantages claimed for the theory of feature geometry is that it provides solutions to certain long-standing problems in the feature theory of SPE (Yip 1989). Consider the following places of articulation, together with their standard SPE classification:

labial	alveolar	palatal	retroflex	velar	uvular
–cor	+cor	+cor	+cor	–cor	–cor
+ant	+ant	–ant	–ant	–ant	–ant

Such an analysis predicts the following natural classes:

[+cor]:        {alveolar, palatal, retroflex}  
[–cor]:        \*{labial, velar, uvular}

But while the [+cor] class is frequently attested in phonological rules, the [–cor] class is never found. The problem here is that the feature theory of SPE embodies an implicit claim that, if one value of a feature denotes a natural class, then so will the opposite value. This is hard-wired into the theory: it is impossible to give oneself the ability to say [+F] without simultaneously giving oneself the ability to say [–F]: equivalently, it is impossible to give oneself to refer to a class of segments without simultaneously giving oneself the ability to refer to the complement of that class.

Sagey (1986) suggested replacing the standard classification with one based on active articulators:

labial	alveolar	palatal	retroflex	velar	uvular
LAB	COR	COR	COR	DOR	DOR

Such a theory predicts the following classes:



LABIAL:	{labial}
CORONAL:	{alveolar, palatal, retroflex}
DORSAL:	{velar, uvular}

Under this approach, the problematic class mentioned above simply cannot be mentioned: the desired result. Unlike the standard theory, which is strictly BIVALENT, this proposal allows for MONOVALENT feature specifications, which have no 'automatic complement'. An essential implicit adjunct to this proposal is the notion that one cannot explicitly refer to the class of features which *lack* the coronal specification (referred to by Clements (1987) as 'invisibility of zero'): if we allow ourselves this capacity, then the feature coronal is not monovalent at all, but effectively binary.<sup>1</sup>

This, then, is the first problem for which feature geometry is claimed to provide a solution. Consider now the same argument, but with respect to the feature [anterior]; we predict the following classes:

[+ant]:	*{labial, alveolar}
[-ant]:	*{palatal, retroflex, velar, uvular}

Here the problem is even worse. As commonly remarked in the literature (e.g. Kenstowicz and Kisseberth 1979:248) there seems to be no phonological process for which this feature denotes a natural class in its own right. [Anterior] is effectively an ancillary feature, whose function is the subclassification of [+coronal] segments, not the cross-classification of the entire consonant inventory. Sagey (1986) again exploited the hierarchical structure provided by feature geometry to express this ancillary status, making the feature [anterior] subordinate to the CORONAL node:

---

<sup>1</sup>At least in the classificatory sense with which we are concerned here. There is another aspect to monovalency, the 'combinatoric' aspect prominent in so-called 'particle' approaches. These two distinct dimensions of monovalency are often confused. A number of theories, such as Dependency Phonology, which claim to have exclusively monovalent features, also give themselves the capacity to refer to complement sets, through the use of a negation operator. Note too, that while some (most) theorists (correctly, in our view) maintain the non-referable interpretation of the absence of monovalent features (Archangeli and Pulleyblank in press; Itô and Mester 1989), others, such as Paradis and Prunet (1991) explicitly allow 'reference to a gap', citing in support Kiparsky (1985). Thus some indeterminacy surrounds this issue in current phonological theory.

labial	alveolar	palatal	retroflex	velar	uvular
LAB	COR   +ant	COR   -ant	COR   -ant	DOR	DOR

As mentioned previously, the theory of structured specification provides a richer framework of classification than that supplied by SPE. Let us now consider the picture that emerges from the CLASSIFICATION HIERARCHY that results from this approach (introduced in Chapter 5). Consider, for example, the following set of fricatives:

	f	s	ʃ	x
COR	-	+	+	-
ant		+	-	

Note here that the absence of a specification in a cell of the table denotes that the segment is *undefined* for that feature, not that the segment has a redundant '±' specification for that feature. Thus we represent /f/ and /x/ as undefined for the feature [anterior], and we assume first, for the purposes of exposition only, that coronal is a binary feature. The classes associated with these features are as follows:

$$[+cor] = \{s, ʃ\}$$

$$[-cor] = \{f, x\}$$

$$[+ant] = \{s\}$$

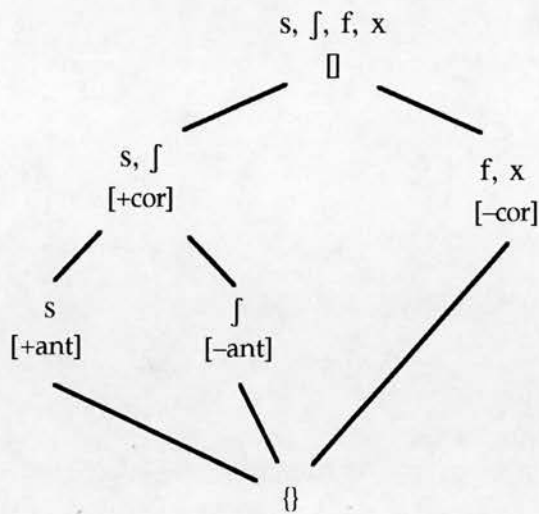
$$[-ant] = \{\}$$

$$\begin{bmatrix} +cor \\ +ant \end{bmatrix} = \{s\}$$

$$\begin{bmatrix} +cor \\ -ant \end{bmatrix} = \{\}$$

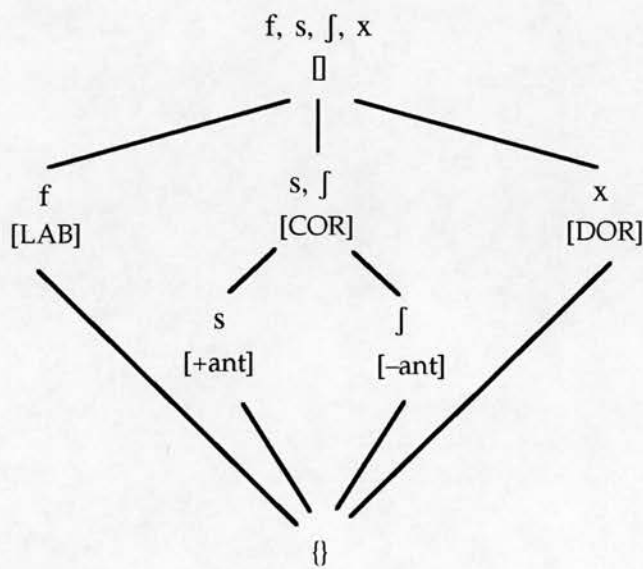
The specifications form the following lattice structure:

(1)



Note that the notion of subclassification is naturally expressed in the framework: the features  $[\pm\text{anterior}]$  subclassify the class denoted by  $[+\text{coronal}]$ . Further, relaxing the binary requirement for gross place specifications, and employing monovalent (non-complemented) nodes, we obtain the following lattice:

(2)



Note that the classes made available by this theory are just those made available in Sagey's reconstruction of feature geometry.

The representation of the segment /s/ will thus be:

(3)



It will be noted that the lattice-theoretic structure is here isomorphic to the structure advocated by feature geometry.

At first sight, it may seem surprising that we are able to subsume aspects of feature geometry within specification theory — a reading of recent phonological literature tends to give the impression that ‘feature geometry’ and ‘underspecification theory’ are independent modules. As we have seen, however, specification theory subsumes just those specific innovations to feature geometry suggested by Sagey (1986), and both of Sagey’s suggested innovations are concerned to remedy defects in the apparatus of *classification* employed by the SPE framework: a framework which is exclusively BIVALENT and CROSS-CLASSIFICATORY. The framework of structured specification employs a richer classificatory apparatus than SPE, one which is wedded to neither of these principles.

Further, there is an acknowledgment in the literature — and, moreover, from the primary underspecification theorists — that there is indeed an intimate connection between the concerns of feature geometry and specification theory. Thus Steriade (1987), the fullest statement of the principles of so-called ‘contrastive specification’, in a section titled ‘Types of specification’ identifies as her first type TRIVIAL UNDERSPECIFICATION:

A segment may lack specifications for a feature value at all stages in the derivation: for instance one can argue that labials never acquire specifications for the feature [anterior], because anteriority is a feature characterizing exclusively the position of the tongue blade, an articulator that is not active in the

production of a labial. In this case, labials are trivially underspecified for [anterior]

And Archangeli (1988) refers to this same phenomenon as INHERENT UNDERSPECIFICATION. As Steriade (1987:340) notes:

My choice of terms is not meant to suggest that trivial underspecification is a trivial aspect of the theory. On the contrary [...] we do not know at present which sounds and which features stand in the relation of trivial underspecification, whether the relation is universal or language specific, or what if anything the difference between trivial and non-trivial underspecification follows from.

The structured specification approach expresses this intimate connection, and provides a framework within to investigate the unresolved issues mentioned by Steriade.

The lattice in (3) above also provides a concrete example of the claim that the structured specification approach is in no way confined to binary features. As discussed above with reference to the non-Boolean properties of classification, our approach treats binarity as a secondary, contingent property of features: not an inherent property, but a function of the set of segments within which the feature operates, and of the features which are used in the classification. Note, too, that subclassificatory hierarchy of the sort we have exhibited is not confined to monovalent (in our terms, non-complemented) features: that is, there is nothing to stop the dominating node being [+coronal], i.e. a node which feature geometry would regard as a genuinely binary feature. In our approach, the representation of /s/, say, remains the same whether the node labelled CORONAL is disjoint with some other node (has a lattice-theoretic complement) or not. Note, too, that in our approach it is a natural result that one and the same feature may be monovalent in one language, but bivalent in another — the difference being a function of whether the complement class to the class picked out by the feature is an element of the lattice. This proposal is in line with a number of current proposals (Archangeli 1988:205).

## **7.2. First order and second order terms**

We have seen an isomorphism between the hierarchical organization of ARTICULATOR NODES and their dependents, and the hierarchical organization of structured specification. Does this isomorphism extend to so-called CATEGORY or CLASS NODES of feature geometry and their dependents? We

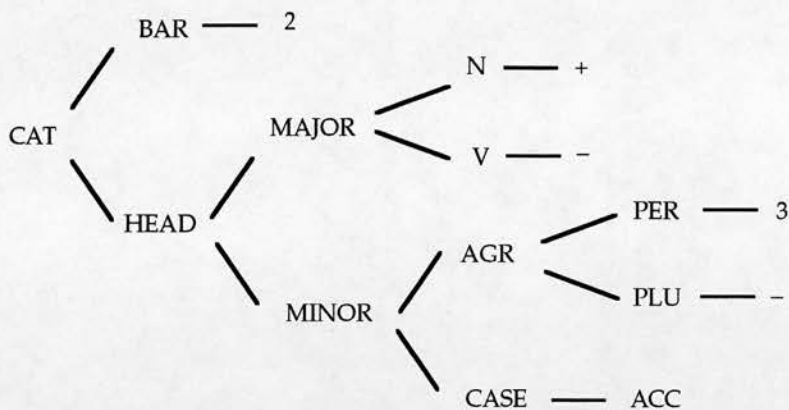


claim that it does not; and further, the fact that it does not is a consequence of the fact that the relation between class node and dependents on the one hand, and articulator nodes and dependents on the other, is fundamentally different in nature. The difference between these two types of node was well-documented and formally unproblematic in early feature geometry. However, the innovations to feature geometry discussed in the previous section — introduced by Sagey (1986) and adopted by virtually all practitioners in the field — have led to a blurring of the distinction. Feature geometry so extended is powerless to express this fundamental difference formally, this formal shortcoming has led to conceptual confusion.

A characterization like PLACE is fundamentally different in kind from a characterization like CORONAL. CORONAL is an attribute of objects (segments); PLACE is an attribute of features, not segments. This fact is reflected in the everyday speech of phonology, where we talk of ‘coronal segments’, but ‘place features’. Just as features like CORONAL allow us to refer to natural classes of segments, terms like PLACE allow us to refer to ‘natural classes of features’ (Clements 1987), features which behave as a unit in various phonological rules. ‘Place feature’ thus has the sense of ‘head feature’ in unification-based syntactic theory. Consider, for example, the following discussion of head features in syntax (Gazdar et al. 1985:21–22):

One particularly useful elaboration of the theory of features, of which we adopt a special version below, involves letting features take other feature specifications, or whole categories, as their values, an idea suggested, in effect, by Anderson (1977, 1981), Kay (1979), Evans (1982), and Pollard (1982).

In Gazdar and Pullum (1982), a theory of features was developed along these lines, using graph theory to provide the basic concepts [...] The advantage Gazdar and Pullum derived from the graph theoretic approach was that whole clusters of feature specifications could be picked out if they shared a mother node in the graph. [...]



HEAD, by virtue of its position in the graph, provides a way of pulling out a cluster of related feature specifications that must be referred to for a particular purpose

Gazdar et al. (1985:23) employ a set theoretic reformulation of this position:

we still want to be able to refer to collections of feature specifications such as those dominated by HEAD and MAJOR [...] The way we shall do this is by associating names such as **HEAD** with designated subsets of the set of features. In fact, we define **HEAD** as follows:

**HEAD** = {N, V, PLU, PER, VFORM, SUBJ, PFORM, AUX, INV, PAST, PRD, ADV, SLASH, AGR, SUBCAT, BAR, LOC}

This reformulation brings out clearly the notion that HEAD classifies — picks out sets of — features, in the same way that those features classify objects.

The original formulation of feature geometry insists on precisely this distinction. Clements (1985:228) writes:

In this conception, individual features are organized under hierarchically superordinate nodes, which I will term **CLASS NODES**. The class nodes themselves are dominated by a yet higher-level class node, which (following Mohanan) I will term the **ROOT NODE**.

Clements (1985:241) also provides an informal set theoretic perspective on this organization:

We have so far said nothing about the phonetic content of the various class tiers [...] The class of laryngeal features comprises the set [spread], [constricted], and [voiced]

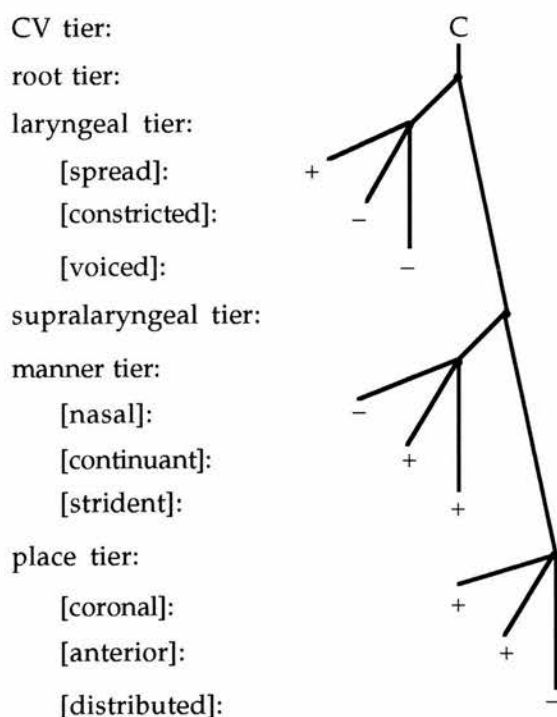
or, as we might say, we can define **LARYNGEAL** as follows:

**LARYNGEAL** = {SPREAD, CONSTRICTED, VOICED}

And, following Clements (1985:248), we may define **PLACE** as:

PLACE = {CORONAL, ANTERIOR, DISTRIBUTED}

Clements, then, carefully distinguishes organizational nodes (class nodes) from phonetic content nodes. He provides for organizational nodes to dominate content nodes, and for organizational nodes to dominate organizational nodes; what he does *not* provide for is for content nodes to dominate content nodes. In Clements (1985) [coronal] is still a binary feature, as are all content features; and such features are always terminal in Clements's diagrams, and never dominate other features. The following is representative, a partial representation of the segment /s/ (Clements 1985:248):



As we have seen, however, Sagey (1986) introduced the possibility that content nodes may dominate content nodes; and that content nodes may be monovalent. Under this proposal, which is now standard, the corresponding place organization for /s/ is (partially):

place tier:  
 coronal tier:  
     [anterior]:  
     [distributed]:



CORONAL, then, is a special type of feature: a PLACE feature. The problem, and the potential confusion, lies in the fact that, in another sense of 'type', [+anterior] is a type of CORONAL feature.

The fact that there are two different relations involved here is reflected in the following stipulation by Archangeli and Pulleyblank (in press:6):

We follow Sagey in allowing such unary class nodes [as Coronal] to be terminal. [...] Unlike work such as Sagey (1986) [...] we postulate that all terminal nodes must correspond to phonetic content. Strictly organizational nodes, such as Place and Laryngeal nodes, may not occur terminally since they have no phonetic content'.

The reference to CORONAL as a 'class node' is symptomatic of the confusion that exists regarding this issue in current phonology. Yet, despite informally blurring the distinction between content node and class node in their terminology, Archangeli and Pulleyblank explicitly invoke this distinction, and claim that the two sorts of node have a different status in the phonology. But note that this distinction, a fundamental one concerning the domain of classification — objects or features — is not reflected in the *phonological formalism* itself. In current feature geometry, there is no way to tell from the representation what the 'type' of the relation between dominant and dominated feature is. Thus the following fragment of a complete representation encodes both types of relation —  $R_1(\text{PLACE}, \text{CORONAL})$  and  $R_2(\text{CORONAL}, [+ant])$  — in the same way, immediate dominance:

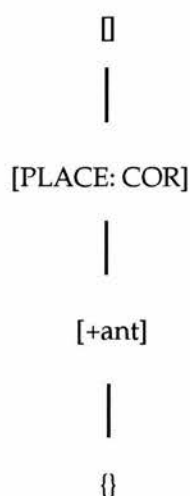


(The distinction is not captured by the fact that [+ant] is terminal, since advocates of this approach explicitly allow for nodes like CORONAL to be terminal.) We might say that feature geometry currently exhibits a ‘confusion of orders’: CORONAL is a first order attribute (classifying objects); PLACE is a second order attribute (classifying classifiers). The problem for feature geometry is, then, that [+anterior] is also a first order attribute, and one which, moreover, is hierarchically organized with respect to CORONAL. Feature geometry uses the same formal device — immediate dominance — for both types of relation; and thus the entire hierarchy of feature geometry confuses two quite distinct orders of classification: the classification of segments and the classification of features.

CORONAL then, is a type of PLACE feature ‘directly’ as it were, by definition. [+anterior] is not a type of CORONAL feature in the same direct sense: rather, [+anterior] *segments* are a type (subclass) of CORONAL *segments*. By virtue of this fact, there is clearly an implicational relationship between the features CORONAL and [anterior], and this is what the theory of structured specification reflects. Note how we determine the relationship: we identify the CORONAL segments, we identify the [+anterior] segments, and we note that one is a subclass of the other. On the contrary, however, we cannot identify the ‘PLACE segments’. We *can* identify all the segments which have a PLACE specification, but all this means is that we find the union of CORONAL and LABIAL and DORSAL segments: segments which have specifications which are PLACE specifications. This is simply a reflex of the fact that certain nodes have ‘phonetic content’: it is this property that allows them to identify classes of segments. However, due to the shortcomings of the notation of feature geometry, Archangeli and Pulleyblank are forced to *stipulate* that only representations which include such features (and not simply contentless, organizational features) constitute well-formed representations.



In the theory of structured specification, the distinction is explicitly reflected: diagrammatically, the two classifications are orthogonal:



That is, the nodes of the lattice exhibit internal attribute-value structure: structure of the kind that characterizes the original formulations of feature geometry. We will elaborate on this aspect of node internal structure when we examine sequence structure constraints in the next chapter.

Thus our position is that the phenomena addressed in Sagey's (1986) thesis — hierarchy among content nodes, and issues of valency — are properly formalized in terms of the specification hierarchy, and not the hierarchy of feature geometry at all. The 'dominance' relation in the two cases has entirely different interpretations. The hierarchical organization which formed the focus of Sagey's thesis is, as it were, in 'another dimension', and this second hierarchical dimension is in fact the one provided by the specification (dually, classification) hierarchy. It is thus just *these* aspects of feature geometry which are subsumed by specification theory, not the original aspects of the theory as advocated by Mohanan and Clements.

The confusion that results from the mixing of orders is apparent in the following remark by Paradis and Prunet (1991:23) repeated from the previous chapter:

Reference to [0F] creates potential ternary distinctions only if [F] is a binary feature [...] However, class nodes, such as articulators or PNs, are unary

Questions regarding the distinctness of representations are impossible to resolve when the semantics of the formalism is disregarded in this way,

treating content nodes and organizational nodes as a natural class. The confusion is compounded by certain approaches to underspecification theory, which propose to underspecify nodes without regard to whether they are first order or second order. Compare the following summary by Paradis and Prunet (1991:6):

most contributors to this volume argue that the Coronal articulator is the unmarked (predictable) articulator. [...] not only do coronals lack a Coronal articulator but also, for many authors in this volume, they lack a PN (Place Node).

We return to this issue when we examine underspecification theory.

### **7.3. Redundancy of articulator nodes**

In this section we present a novel argument against encoding redundancy by omission: the argument focusses on a problem which arises when we try to combine the requirement to omit redundant information with the independent assumptions of feature geometry.

[+Anterior] and [–anterior] segments are predictably CORONAL. It follows that if we are to take seriously the requirement to omit predictable information from underlying representations, then we should omit the CORONAL specification from segments which are distinctively [±anterior] — after all, in the inverse case, if a segment is the only CORONAL segment in the inventory, the fact that it is [±anterior] is thus redundant, and is expressed precisely by omitting the [anterior] specification from underlying representation, retaining simply CORONAL as the terminal node. However, when we combine the requirement to omit the (redundant) CORONAL node, with the hierarchical organization of feature geometry, we encounter a serious problem. The result is a ‘geometrically discontinuous’ segment, with [±anterior] specified at the periphery, but with the supporting structure — which is redundant — omitted. Thus an inconsistency arises when the proposals of feature geometry are combined with the proposal of non-redundant specification.

The theory of structured specification — in which redundant information is retained — does not suffer from this problem, however. Thus the lattice-structure exhibited in (2) above is not simply a notational variant of feature geometry: it preserves the desirable aspects of hierarchical feature

organization, but without the theoretical inconsistencies inherent in the standard model.<sup>2</sup>

Note, too, that the problem of redundant articulator nodes is only a substantive problem for a theory in which content nodes are hierarchically organized. That is, a node like PLACE is not redundant given the specification CORONAL: there is no implicational relationship between these two nodes, since they are taken from different orders of classification. However, there *is* an implicational relationship between the content nodes CORONAL and [ $\pm$ anterior], and thus they properly fall under any requirement to omit predictable information.

#### 7.4. 'Uncharacterized' segments

Another virtue claimed for the theory of feature geometry is that within it one is able to employ 'uncharacterized' representations, which are thus able to express the notion of a 'defective' segment (Lass 1976). The classic such case is that of segments such as [h ?] which are uncharacterized for any place features. This is a simple and elegant idea: and clearly, in such cases, one of the defining properties of such an element is precisely the lack of specifications. However, when this notion is combined with a theory of underspecification, difficult problems arise. The reason is simple: the same formal mechanism — the lack of a specification — is forced to do two different things: it represents 'unspecifiable' status on the one hand, and redundant status, on the other. For any given representation which lacks certain features we cannot tell which interpretation is appropriate.

If, on the other hand, we know that every representation is fully-specified, as we do in the theory of structured specification, then the ambiguity simply disappears. In such a model, the lack of specification genuinely means that the segment is undefined for that property — not that it *has* such a property, but redundantly, or by default. In other words, the use of the mechanism of blanks to encode redundancy and defaults preempts the use of this mechanism for cases where there is either (i) a lack of information, which can be supplied by alternative contexts (archisegments); or (ii) an inapplicability

---

<sup>2</sup> To our knowledge, this inconsistency has not been discussed by advocates of underspecification, but the same point has recently been made by Pierrehumbert (1992).

of information, the 'undefined' case (defective segments). Structured specification avoids such problems.

These issues, however, involve not just feature geometry, but the interaction of feature geometry with underspecification theory. Accordingly, we delay further consideration of these matters until we have explicitly compared our approach with the assumptions of underspecification theory.

## 8. Structured specification and the lexicon

---

### 8.1. Introduction

In this chapter we discuss the relevance of our approach to a number of issues in lexical representation and recognition. We first extend the principles of structured specification to the wider syntagmatic domain, and provide an account of SEQUENCE STRUCTURE CONSTRAINTS along the same lines as our previous treatment of segment structure constraints. Extending the approach in this way results in a lexical organization of phonological structure which is in many ways parallel to the organization of the lexicon in unification-based approaches to syntax and semantics, a key element of which is the notion of SHARED STRUCTURE. The significance of this result is that, despite our decision to include all redundant information in lexical entries, we are still able to provide a maximally efficient encoding of lexical representation. Under this approach, redundant structure is shared structure. The phonological lexicon takes the form of an INHERITANCE HIERARCHY, with individual morphemes occupying the terminal nodes. The resulting hierarchy has close affinities with the network organization proposed by Waterson (1981) and can be seen as a natural extension of that approach. We *identify* phonotactic constraints with the structure of this lattice of representations, in exactly the same way as we previously identified segment-structure constraints with the interrelationships between features in the lattice of segments.

Inheritance hierarchies have been the focus of attention in the field of knowledge representation for some time, and we take advantage of this research to provide a formal account of defaults in terms of NONMONOTONIC INHERITANCE.

Our approach to phonological redundancy and phonotactic constraints provides a bridge from phonology to the model of lexical representation employed by Zue and his colleagues in speech recognition. In addition, a consequence of structured specification approach to reciprocal dependency is that the phonology is agnostic with respect to which of the mutually dependent properties should be regarded as predictable, and which unpredictable. Artificial distortions introduced in the standard model are



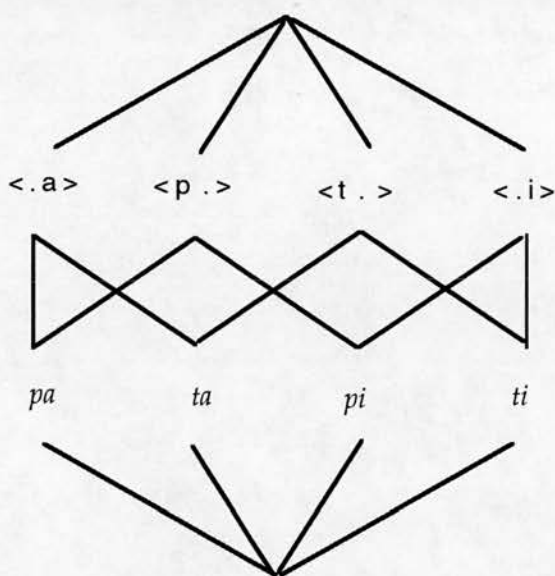
avoided. Rather we provide a 'grammar of redundancy' which is neutral with respect to the processing demands that are placed upon it. In the final section of this chapter, we consider the implications of this aspect of the model, and show that it is compatible with recent results in speech perception.

## 8.2. Sequence structure constraints

Consider the language consisting of the strings:

{<pa>, <ta>, <pi>, <ti>}

These strings can be arranged in a lattice as follows:



Now consider the effect of a gap in the inventory of syllable types:

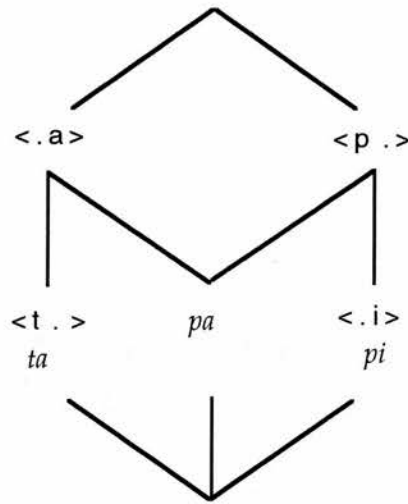
{<pa>, <ta>, <pi>, \*<ti>}

Given this, the following implicational constraints hold:

<t.>  $\supset$  <.a>

<.i>  $\supset$  <p.>

which can be represented in the following lattice:



It is clear that the vowel is distinctive for <pi>, whereas the consonant is distinctive for <ta>.

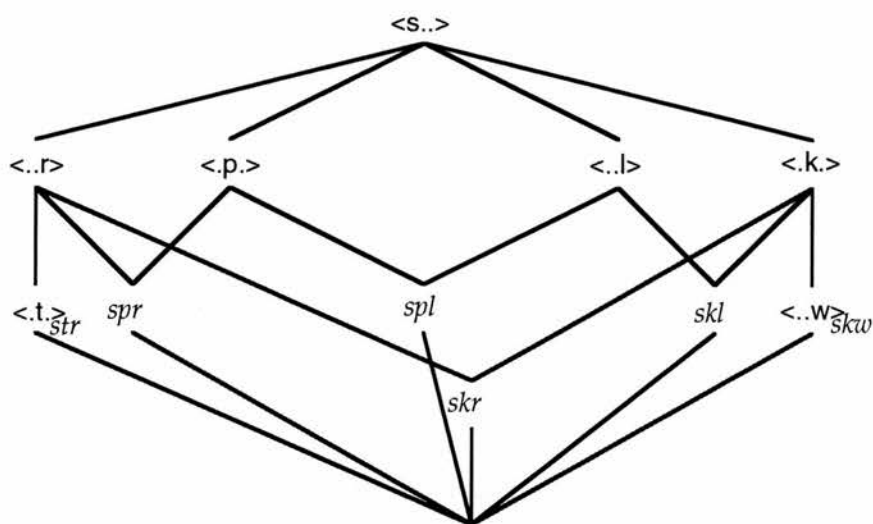
In an approach in which redundant information is omitted from underlying representation, the lexical representation of *pi* would be:

<.i>

In our approach, on the other hand, the redundant information is explicitly represented, but coded as redundant by its position in the lattice:

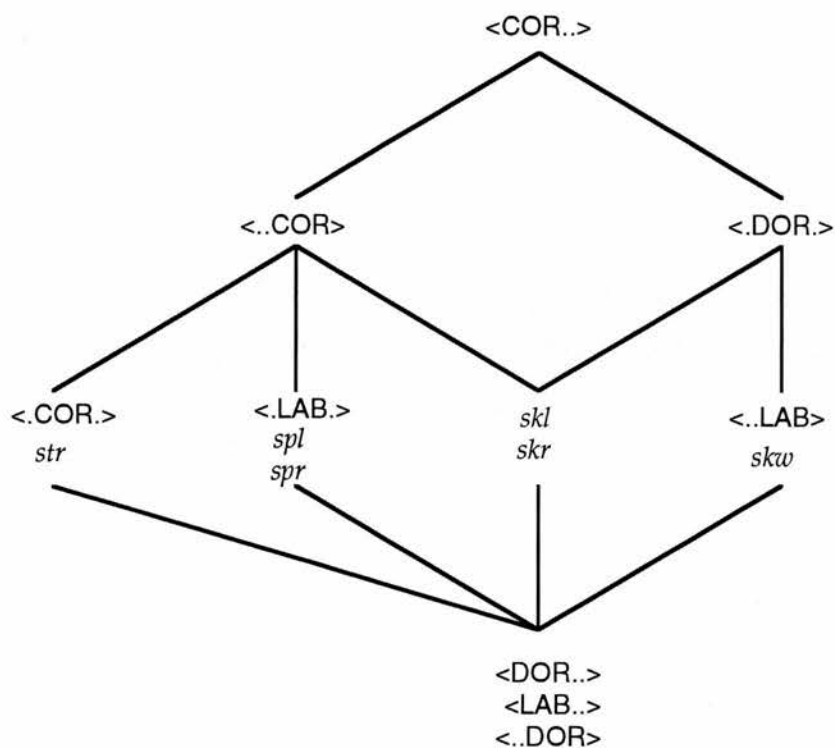


To take a more realistic example, the sequence structure constraints obtaining in initial CCC-clusters in English are coded in the following lattice:

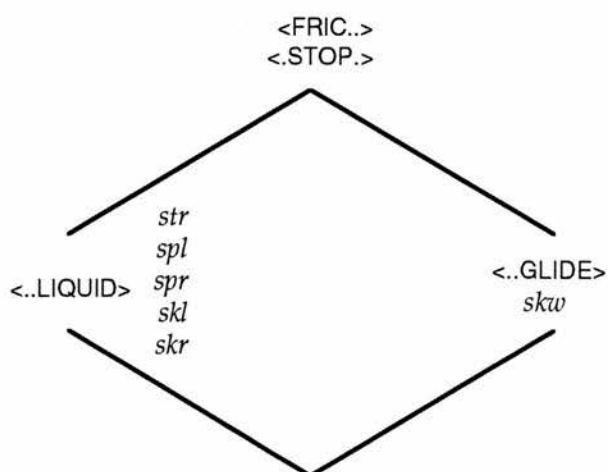


The fact that all such clusters are /s/-initial is directly expressed by the position of this node at the top of the lattice. Two other implicational relationships are encoded: all clusters containing /w/ have a preceding /k/; all clusters containing /t/ have a succeeding /r/.

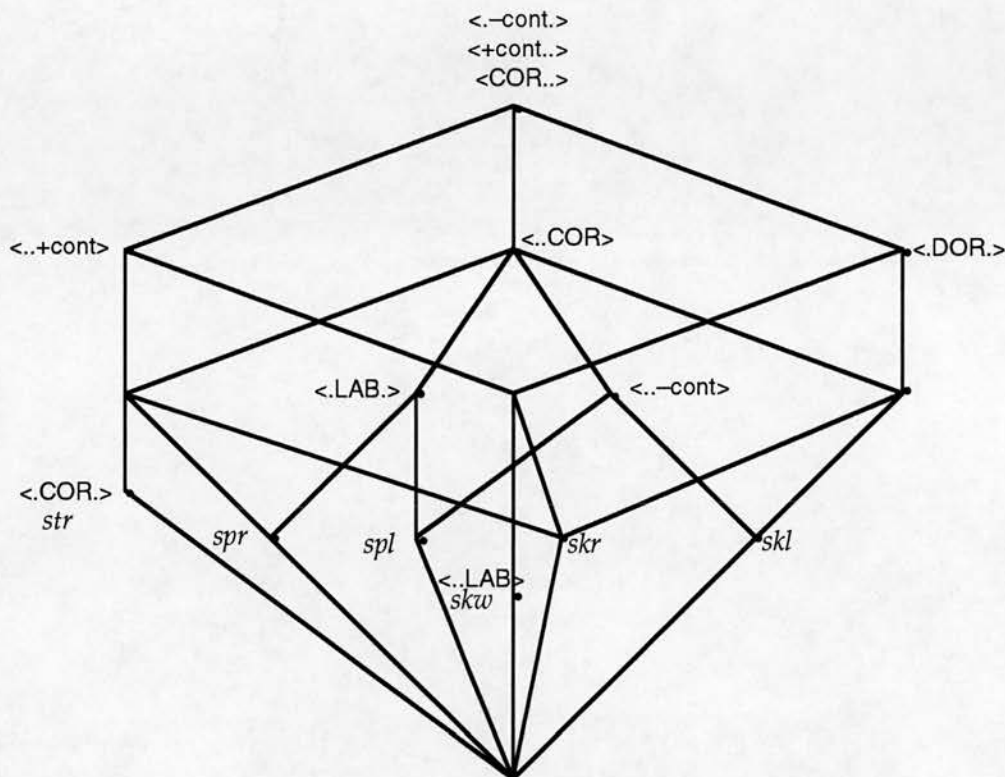
Of course, we are not confined to using unanalyzable symbols like /s/ and /r/: we can combine featural decomposition and the lattice representation of *segment* structure constraints into this sequence structure lattice. In the following diagram we have replaced phoneme symbols with just the place feature of each segment:



Clearly, place information alone is insufficient to distinguish all clusters, but it is perhaps surprising how great is the functional load which even such gross place specifications as LABIAL, CORONAL and DORSAL bear here: they are not sufficient, but are nearly so. Gross manner specifications, on the other hand, have virtually no distinctive function in such clusters, since all initials are fricatives, all medials are stops, and all finals glides/liquids:



If we take /l/ to be non-continuant, the addition of this feature to the place lattice will be sufficient to distinguish all clusters, since in order to do so we need only to distinguish /l/ and /r/ in final position. The result may be diagrammed as follows:



Of course, once the remainder of the word to which these clusters are prefixed is taken into account, it may be redundant to distinguish among /l/-final and /r/-final clusters in terms of some extra feature. Perhaps, for example, /spl/ and /spr/ words can be independently distinguished with regard to vowel quality; perhaps /skl/ and /skr/ words can be distinguished by reference to prosodic structure. To the extent that this is so, of course, the redundancy will be encoded in the structure of the completed lattice. We return to this point when we consider the implications of the approach for word-recognition in section 8.8.

What is the interpretation of a symbol like <s..>? In effect, it represents the combination of properties

$$\exists x: s(x) \ \& \ \text{initial}(x).$$



That is, positional information is coded into the lattice by representing positions as properties. The lattice above thus implicitly employs the properties 'initial' 'medial' and 'final'. This ordering information is implicit in the notation used for strings. Strings are ordered sequences of symbols drawn from some fixed alphabet. If we were to make the ordering explicit, we could do so either by defining a linear precedence relation  $LP$  over the symbols:<sup>1</sup>

$$LP = \{ \langle s, p \rangle, \langle p, r \rangle \}$$

or by placing the symbols in one to one correspondence with some canonically ordered set, such as the set of positive integers, as follows:

$$1 \leftrightarrow s$$

$$2 \leftrightarrow p$$

$$3 \leftrightarrow r$$

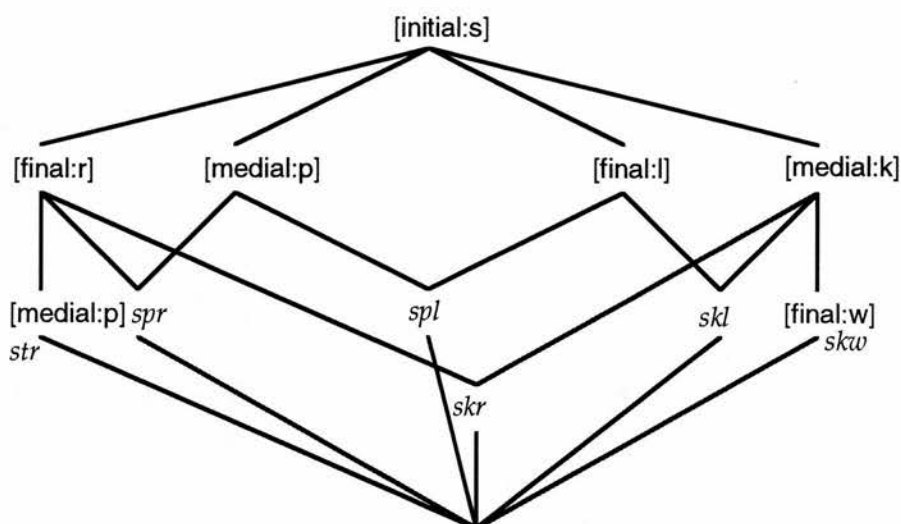
We can incorporate ordering information into the lattice using the machinery of attribute-value structures outlined in Chapter 3. Segments are represented as the values of positional attributes, which are subject to general ordering constraints, thus:

$$\begin{bmatrix} \text{initial} = s \\ \text{medial} = p \\ \text{final} = r \end{bmatrix}$$

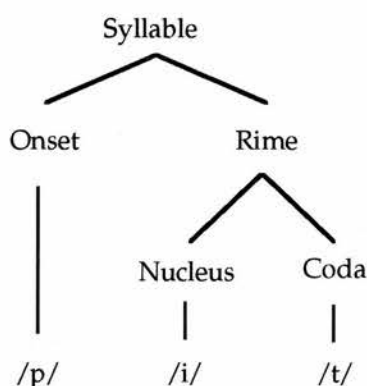
The lattice thus becomes:

---

<sup>1</sup>See Coleman (1991) for an illuminating account of phonological and phonetic aspects of order in a unification-based framework.



Syllable structure can be represented in attribute value terms in a similar way. The structure:



Can be expressed in the following attribute-value matrix:<sup>2</sup>

$$\left[ \text{syllable} = \left[ \begin{array}{l} \text{onset} = p \\ \text{rime} = \left[ \begin{array}{l} \text{nucleus} = i \\ \text{coda} = t \end{array} \right] \end{array} \right] \right]$$

By constructing the lattice in terms of such attribute-value matrices, rather than those associated with simple strings, the hierarchy can be extended to directly encode syllable structure constraints.

---

<sup>2</sup>See Coleman (1991) for an extensive treatment of English syllable structure in these terms.

### 8.3. The hierarchical lexicon

We now return to the notion of extending unification-based approaches to syntax to the sound domain. Unification-based grammars preserve the distinction between syntax and the lexicon. Put simply, the lexicon contains 'syntactic atoms': objects with no internal syntactic structure, but which are organized into larger structures according to syntactic rules. This is not to say, however, that these lexical objects have no internal structure at all. Indeed, many grammars developed within the unification framework employed lexically oriented analyses: the very names 'Lexical Functional Grammar' and 'Head-driven Phrase Structure Grammar' reveal as much, and this lexicalist approach is also evident in the extensions utilizing categorial grammar. Accordingly, these approaches developed increasingly complex lexical entries, simplifying syntactic rules with a concomitant elaboration of the lexicon. Such an entry is exemplified by the following representation of the verb 'storms' from Shieber (1986:35):

$$\left[ \begin{array}{l} \text{cat} = V \\ \text{head} = \left[ \begin{array}{l} \text{form} = \text{finite} \\ \text{trans} = \left[ \begin{array}{l} \text{pred} = \text{storm} \\ \text{arg1} = \langle 2 \rangle [ ] \\ \text{arg2} = \langle 1 \rangle [ ] \end{array} \right] \end{array} \right] \\ \text{subcat} = \left[ \begin{array}{l} \text{first} = \left[ \begin{array}{l} \text{cat} = NP \\ \text{head} = \left[ \text{trans} = \langle 1 \rangle \right] \end{array} \right] \\ \text{rest} = \left[ \begin{array}{l} \text{first} = \left[ \begin{array}{l} \text{cat} = NP \\ \text{head} = \left[ \begin{array}{l} \text{agreement} = \left[ \begin{array}{l} \text{number} = \text{singular} \\ \text{person} = \text{third} \end{array} \right] \\ \text{trans} = \langle 2 \rangle \end{array} \right] \end{array} \right] \\ \text{rest} = \text{end} \end{array} \right] \end{array} \right] \end{array} \right] \end{array} \right]$$

As Shieber notes, although the grammars which manipulate these representations are thereby able to remain relatively simple, a new problem arises in dealing with such unwieldy lexical feature structures: 'Clearly, no one is willing to write such complex and redundant feature structures for each lexical entry'. Accordingly, techniques were sought for expressing lexical entries in a compact notation. As we will see below, the techniques employed were not newly invented for the purpose, but borrowed from approaches to knowledge representation in artificial intelligence.

The phonological parallel to a 'syntactic atom' is a 'phonological atom'. However, there is a fundamental difference between the atoms of phonology and the atoms of syntax: the atoms of phonology have syntagmatic structure. That is just to say we don't store segments, but morphemes. But morphemes, while syntagmatically atomic from the point of view of syntax, are syntagmatically complex from the point of view of phonology. From the point of view of syntax, syntagmatic structure is grammatical (as opposed to lexical) structure: that is, it is not stored but 'built'. But from the point of view of phonology, STORED SYNTAGMS are in fact the norm.<sup>3</sup> Thus while it is appropriate to consider approaches to complexity and redundancy in lexical storage developed in unification lexicons, such approaches will need to be extended to handle syntagmatic structure when translated to the sound domain: to handle what traditionally has been referred to as 'sequence structure' redundancy in addition to 'segment structure' redundancy.

There is another parallel with the development of syntactic theory here. The approaches to efficient representation of lexical information were developed by researchers working on large coverage, computationally explicit treatments of a single language, often working for corporations such as Xerox, Hewlett Packard, IBM, and so on. By contrast, current mainstream generative syntax has primarily been concerned with isolated, specific constructions, notable precisely for their apparent exceptionality to standard theories, and which motivate some particular theoretical modification to the standard model: the so-called 'cute-fact' paradigm. The cute-fact paradigm also informs phonology, and in addition, generative phonology has always been most interested in alternations: indeed, once one assumes as an axiom the

---

<sup>3</sup>Perhaps the closest syntactic analogy would be idioms — stored chunks of syntax.

notion of a unique underlying form, a great deal of interest focusses on the principles of modification which map this unique underlier to the variety of surface forms. This naturally leads to a concomitant reduction in the attention paid to the structure of the base forms (or indeed surface forms) themselves, phonotactically conceived. But from the point of view of word recognition, for example, this is precisely the wrong emphasis. Word recognition systems require large coverage, computationally explicit treatments of a single language. From this point of view, it can be argued that contextual variation, whether allophonic or morphophonemic, can best be handled by being simply ignored (Huttenlocher and Zue 1983:175-6):

The scheme proposed by Shipman and Zue can handle allophonic variations, such as the different realizations of /t/ [in 'tree', 'tea', 'city', 'beaten']. This is because contextual variations tend to affect the detailed acoustic realizations of the phonetic segments, as opposed to the gross manner features used in the broad classes. When accessing the lexicon based on broad phonetic classification, detailed allophonic differences are completely disregarded

The scheme of Shipman and Zue was presented in a paper entitled, appropriately for our argument, 'Properties of large lexicons' (Shipman and Zue 1982). Below, we consider in detail how our approach relates to such a perspective.

To return to our main argument: How, then, was the problem of redundancy approached in unification-based lexicons? Consider the complex lexical entry for 'storms' quoted above.<sup>4</sup> Parts of the information mentioned there will be repeated in many lexical entries. For example, all verbs will contain the specification:

[*cat* = *v*]

In addition, all transitive verbs will contain the following information:

$$\left[ \begin{array}{l} \textit{subcat} = \left[ \begin{array}{l} \textit{first} = [\textit{cat} = \textit{NP}] \\ \textit{rest} = \left[ \begin{array}{l} \textit{first} = [\textit{cat} = \textit{NP}] \\ \textit{rest} = \textit{end} \end{array} \right] \end{array} \right] \end{array} \right]$$

---

<sup>4</sup>The exposition here follows Shieber (1986:55ff).



In PATR-II (Shieber et al. 1983), a unification-based grammar formalism, such subparts of feature matrices could be abstracted, named, and arranged in an inheritance hierarchy like those used in many AI knowledge representation systems. Under this approach, various lexical entries are permitted to *share* the information stored in a particular template. The PATR-II system of templates was the first utilization of the inheritance network principle of organization in lexicons within a unification-based framework; and the approach was elaborated and incorporated into HPSG. This work was developed by a group at Hewlett-Packard, and their use of the HPRL knowledge representation language led directly to an inheritance-based organization of the HPSG lexicon.

An INHERITANCE NETWORK is a labelled directed graph whose nodes represent individuals and classes, and whose links represent various relations between the nodes (Touretzky et al. 1987). The primary relation is the IS-A link, which is written as  $x \rightarrow y$ . A TREE-STRUCTURED inheritance system is one whose nodes and IS-A links form a tree: that is, any node has just one mother. A MULTIPLE-INHERITANCE system on the other hand, allows nodes to have multiple superior nodes. A UNI-POLAR inheritance system has just one sort of link: BI-POLAR systems contain in addition links of form  $x \rightarrow y$ , called IS-NOT-A links. A uni-polar system thus cannot directly express negative statements. There are further classifications of inheritance networks which we will consider later.

It should be clear that the lattices we have developed may be regarded as uni-polar, multiple-inheritance networks. Consequently, we may use techniques developed in knowledge representation languages in the representation of phonological redundancy. However, it is important to note that the representation of phonological information does not *reduce to* general knowledge representation. As the work on underspecification theory makes clear, the status of information on the lower periphery of the network has a peculiar status in phonological representation. This follows from the nature of the objects about which we make deductions in phonology, as opposed to AI.

The objects one encounters in inheritance networks are typically real world, physical objects.<sup>5</sup> Phonology, on the other hand, involves reasoning about objects which are themselves signs, rather than simply denotata. In such a system, issues of redundancy and logical implication take on a new resonance. For example, in certain cases, if a phonological object has a property P, and that property is redundant — implied by some other property of the object — then the object is treated as if it *lacks* the property. In other words, in certain cases reasoning about objects is restricted to properties which are ‘significant’, in the sense that they are unpredictable or ‘distinctive’ in linguistic terms. There is a pun to be exploited here: distinctive = significant; redundant = insignificant. In certain phonological contexts, certain features do not *signify* — in either sense, the colloquial, or the Saussurean. Thus deduction in such a system may proceed in ‘purposeful ignorance’, as it were, of the set of possible entailments for a given set of properties. In knowledge representation systems, on the other hand, predictable properties are considered as having the same status as unpredictable properties. In phonology, these phenomena are currently the focus of the subfield of underspecification theory, but there is a rich and as yet untapped dialogue to be explored between this phenomenon and the notion of INFORMATION SYSTEMS, in which the basic idea is that of identifying an object with a set of propositions true of it and *adequate to define it* (Davey and Priestley 1990: 62). Thus the domain of phonology provides a complex and extensive domain within which to explore the properties of multiple inheritance systems, and the kind of logic they support. This point will become even more relevant as we turn to the structure of default specification.

#### 8.4. Default features and nonmonotonic inheritance systems

In a previous chapter, we introduced the notion of hierarchical default specification, and promised to return to this point in the light of parallel notions in knowledge representation. To this point we now turn.

Recall that inheritance networks can be classified as tree-structured systems or multiple-inheritance systems, and as uni-polar or bi-polar. In addition, inheritance networks can be classified as MONOTONIC or NONMONOTONIC. A

---

<sup>5</sup>In fact, examples in the literature typically revolve around subtypes of elephants, molluscs and birds.

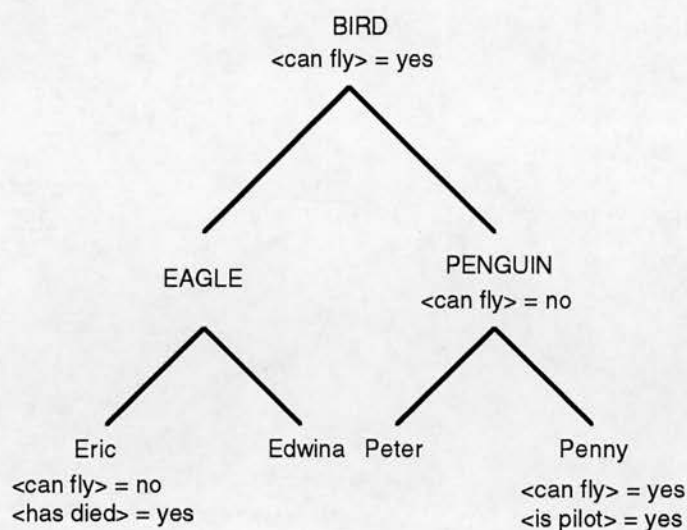
nonmonotonic inheritance system permits exceptions to inherited properties. A number of approaches to incorporating this intuition into inheritance systems have been suggested. One such approach is to add new types of link. Etherington and Reiter (1983), for example, suggest augmenting 'strict' IS-A and IS-NOT-A links with DEFAULT IS-A and DEFAULT IS-NOT-A, together with an explicit EXCEPTION link. Etherington and Reiter supply an formal interpretation of the meaning of such links in terms of default logic (Reiter 1980).

Another approach, and one which was adopted by the HPRL representation language mentioned above, is to employ DEFAULT INHERITANCE. As Touretzky, Horty and Thomason (1987) observe:

The intuition underlying inheritance with exceptions is that claims about subclasses are more specific, and so can incorporate information about exceptional cases. Therefore, subclasses should be allowed to override superclasses.

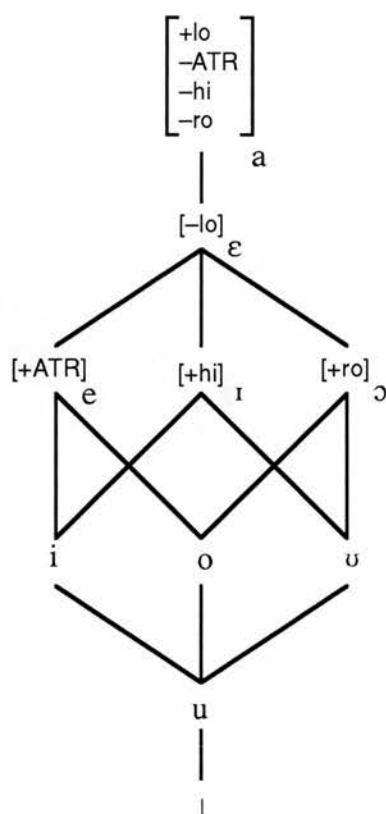
In default inheritance, the default or unexceptional values of a feature are towards the top of the tree, and inherited in the normal way. A daughter node inherits the properties of the parent nodes, unless that property is explicitly contradicted on the daughter itself. In the case of such an explicit contradiction, information lower in the network has precedence over information associated with higher nodes, expressing the intuition above. Thus unmarked features SUBSUME marked features, just as redundant features subsume unpredictable features.

Consider the following analogy in the domain of birds (Gazdar 1990). A default or prototypical bird has the property that it can fly; penguins, however, although they are birds, are exceptional in that they lack this property. Further, even a particular member of a 'normal' class of birds, such as eagles, may be unable to fly for some specific reason; and equally some truly exceptional penguin may in fact be capable of flight. These relations can be set out in the following hierarchy:



The properties associated with Eric, for example, are {<can fly> = yes, <can fly> = no, <has died> = yes}, which contain an apparent contradiction. This contradiction is resolved by giving precedence to the 'more specific' information, the information associated solely with the subclass (Eric) rather than the superclass (birds in general).

Recalling the treatment of default vowel features in Okpe presented in Chapter 5, we may now incorporate unmarked information into the lattice of exceptions, by associating default properties with the top node of the lattice, and interpreting the hierarchy in terms of default inheritance, rather than strict inheritance.



The subsumption relation brings out a deep correlation between archiphonemes and default segments. In the lattice of redundancy, higher nodes on the tree represent archiphonemes, and the lattice of nodes which depend from them represent possible instantiations — a realization space — depending on what further information is ‘unified in’ from other sources. In the lattice of markedness, on the other hand, exemplified above, the down-set of nodes depending from a particular node also constitutes a realization space which determines a range of possible instantiations, depending on what *marked* specifications are added.<sup>6</sup>

Note this difference as well: in the lattice of markedness superordinate nodes are themselves ‘phonemic’ as opposed to ‘archiphonemic’: they require no further specification to be fully realized. Thus default specifications are not added ‘late in the derivation’ as in underspecification theory, but are present in the representation at all times. Accordingly, default specifications are accessible to the phonological derivation, and may be referenced by

---

<sup>6</sup>See Gibbon (1990) for a similar account of defaults using the DATR formalism.



phonological rules. But their status *as* default specifications is explicitly coded by their structural position in the lattice. Just as distinctive information is most specific, and redundant information most general, so marked specifications are most specific, and unmarked specifications most general.

Thus, for example, the node labelled /ɪ/ subsumes /i/ and /u/, realizations which result from unifying marked values, just as an archiphoneme subsumes a set of possible instantiations. However, this node is fully specified as it stands, since without further instantiation it inherits superordinate features.

### **8.5. Combining defaults with redundancy**

We have seen inheritance hierarchies based on redundancy, and inheritance hierarchies based on default specifications: we have taken a lattice-based approach to both distinctiveness and defaults, and seen that subsumption structure informs both these domains. It would seem an obvious next step to combine the two. Precisely this question has been addressed extensively in the literature on knowledge representation. However, as Sandewall (1986:1345) succinctly observes 'Multiple inheritance with exceptions: surprisingly difficult':

Exceptions are fairly easy to deal with in single-inheritance systems, and can be obtained through conventional block structures. Multiple inheritance without exceptions is also easy to deal with theoretically. The combined structure, multiple inheritance with exceptions, offers, however, a number of unpleasant and challenging surprises.

It is easy to grasp the central problem here. We have seen that the redundancy hierarchy represents a network of implications which hold among various properties. Now if one of those properties is superseded by a contradictory specialization, not only that property is affected, but also the network of implications with which that property is implicated.

Early attempts at combining multiple inheritance with exceptions — the FRL system (Roberts and Goldstein 1977) and NETL (Fahlman 1979) — were based on straightforward extensions of tree-structured inheritance systems. However, Touretzky (1986) showed that these approaches were theoretically unsound. Touretzky went on to present a theoretically sound system, but interestingly, other definitions for inheritance were proposed around the

same time (Sandewall 1986, Horty et al. 1987) which, while equally sound and apparently intuitive, did not always agree with the system of Touretzky. In an overview paper, Touretzky et al. (1987:476) conclude:

At the heart of the controversy is a clash of intuitions about certain fundamental issues in inheritance reasoning [...] Just as there are alternative logics, there may be no single 'best' approach to non-monotonic multiple inheritance.

The intuitions in question centered on choices between 'skeptical' vs. 'credulous' reasoning; 'upward' versus downward' reasoning; 'on-path' versus 'off-path' preemption; classical versus intuitive notions of consistency (see Touretzky et al. (1987) for discussion). Investigation of such issues is properly in the domain of workers in knowledge representation, but we believe a specifically phonological contribution can be made to this line of research, in the following terms. Touretzky et al. (1986:481ff.) conclude:

The study of inheritance reasoning has rapidly moved through stages similar to ones that deductive logic has passed through at a much more glacial pace [...] we have described the discovery of several dimensions along which alternative approaches to inheritance reasoning can be generated, all of which appear to be equally sound [...] The existence of these approaches has many theoretical consequences, which we are exploring in our current research. [...] However, this theoretical work needs to be combined with more 'empirical' investigations of the alternatives. Sandewall has suggested that a useful way to approach the question of sound multiple inheritance reasoning with exceptions might be to create a catalog of problematic networks.

As we will see subsequently, the interplay of phonotactic constraints and default information in phonological systems leads to subtle and systematic surface variation. Further, this interplay can be successfully formalized using notions developed in default logic — we present an example of such an account in the final chapter. Phonological systems thus present an interesting empirical domain within which to investigate properties of default behaviour.

## **8.6. The expression of phonotactics**

Halle (1959:30) identifies an important ancillary function of his Condition (5) ('the number of specified features is consistently reduced to a minimum') the condition which we have rejected. The requirement to omit predictable information, not just in the context-free (segment structure) case, but also in *specific* contexts where features are predictable, means that Condition (5)

constitutes 'the device whereby distributional constraints are built into the grammar of the language'. That is, Halle's underlying representations present a pattern of specification — of pluses and minuses on the one hand, and blanks on the other — and this representational pattern is a reflex of the distributional constraints in the language. Having rejected Condition (5), we must examine how it is that such phonotactic constraints are expressed in our approach.

The model of the lexicon which emerges from the structured specification approach makes a novel contribution to the conceptual interpretation of phonotactic constraints. Under our approach, the structural constraints, which in other approaches are expressed by a set of rules or conditions disjoint from the stored forms, constitute the storage pattern itself. That is, rather than the specification pattern being a reflex of the phonotactic constraints, the pattern of specification in a particular representation is identical with the phonotactic constraints. The phonotactic constraints are not stated independently of lexical entries, but are IMMANENT in the structure of lexical entries. Thus we encode the phonotactic constraints directly in the lexical forms, without 'applying' the constraints, or 'mapping' between levels of representation, or 'checking' a list of forms against a list of conditions.

### **8.7. Comparison with Waterson**

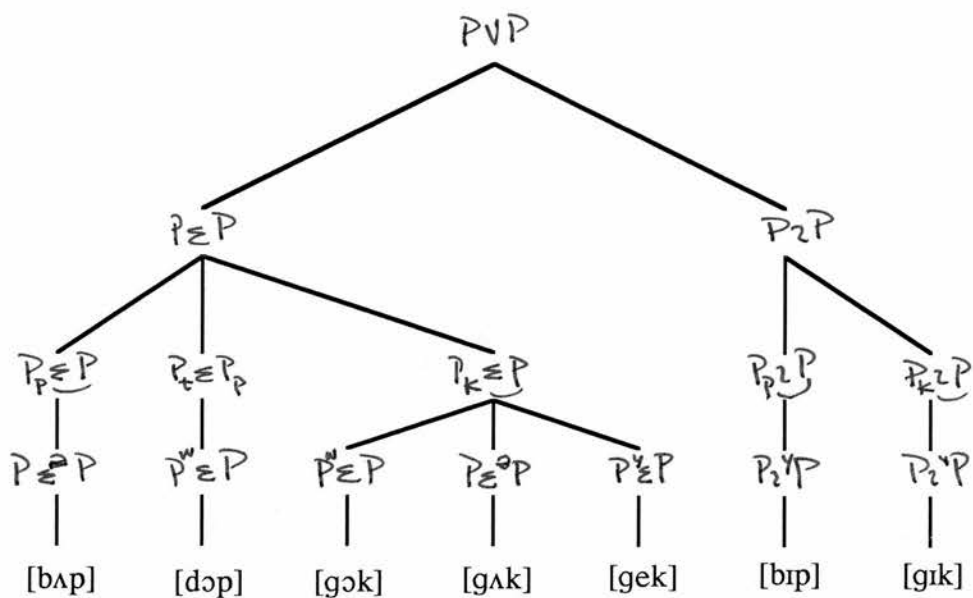
The inheritance networks of knowledge representation are not the only parallel to the approach we advocate. An interesting antecedent is provided by the work of Waterson on child language development (Waterson 1987). Her work is particularly interesting, since it adopts an explicitly Firthian approach to language structure. Recall that according to Anderson (1985:12) the British prosodic school was unique in its reconsideration of the role played by redundancy in phonological representations. Waterson (1981) provides Prosodic Phonology with an explicit model of lexical storage and representation. As we will see, it is strictly comparable to the approach of structured specification. Our aim here is not to provide a detailed comparison of the two approaches, but simply to illustrate the way in which they converge on a similar notion of lexical organization.

Waterson sees her approach as providing an alternative theory of lexical representation particularly appropriate for work in speech perception and interpretation. She provides the following illustration of the organization of

phonological patterns of a child who, at this stage of development, had the following CVC patterns:

- 1) Plosive-vowel-plosive, e.g.: [bʌp] 'cup', [bɪp] 'bib', [gek] 'cake', [gɔk] 'clock', [gɪk] 'stick', [gʌk] 'duck', [dɔp] 'stop'.
- 2) Plosive-vowel-nasal, e.g.: [gɔn] 'gone', [daun] 'down', [dʌn] 'stone'.
- 3) Nasal-vowel-nasal, e.g.: [muum] 'moon', [mæm] 'man'.

Waterson (1981 [1987:112]) selects the plosive-vowel-plosive pattern, which has the largest number of examples, and arranges them in the following 'PVP network':



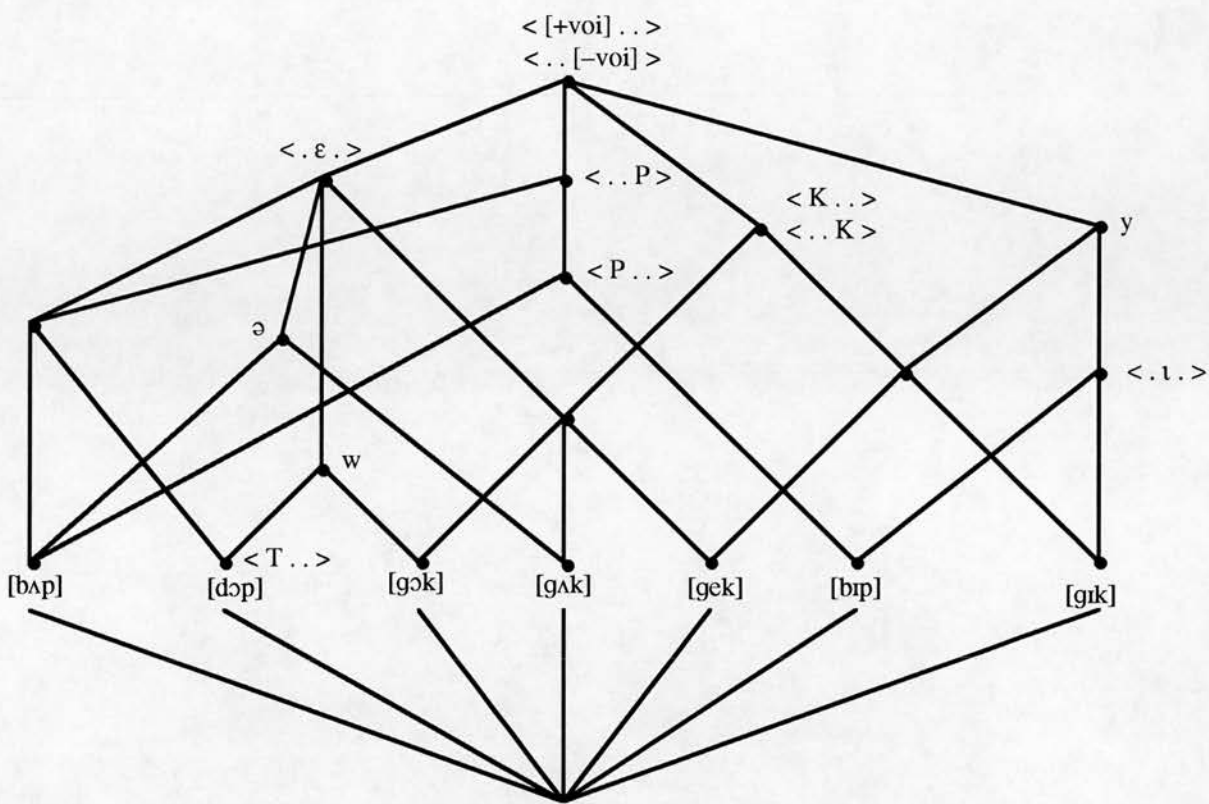
Waterson (1981 [1987:112]) comments:

The symbols α, ε, ι are used to represent the three functional contrasts at the V place of PVP. These function in conjunction with contrasting syllable prosodies which are represented by the symbols y, w, and ə. P stands for the plosive system which has the possibility of a contrast of three terms: p, t, and k.

Waterson's expression of the network as a tree-structured hierarchy has the consequence that certain redundancies cannot be expressed. It also means that arbitrary rankings must be imposed: under the right branch PιP node, for example, the P... P agreement pattern is arbitrarily ranked above the prosody. The problem is that the actual relationship here is one of cross-classification, rather than subclassification. Waterson's network is straightforwardly translated into the following lattice of representations. Note that the 'syllable



prosodies' y, w, and ə classify syllables directly, and are not associated with any particular CVC position:



Unsurprisingly, in the light of our discussion in Chapter 3, the structured specification approach also exhibits clear parallels with systemic classification (Halliday 1961). As noted by Mellish (1988) the possible descriptions allowed by a systemic network are partially ordered by the subsumption relation, and form a lattice. In addition, systemic grammar distinguishes a SELECTION SET, which is simply the set of features extracted in a pass through the network, from a SELECTION EXPRESSION 'which indicates how the various terms in the selections are related in the network' (Henrici 1981: 77). In addition to the actual features, selection expressions contain brackettings and diacritics indicating subclassification, cross-classification and mutual implication which indicate the relationships which the features bear to each other. The structure of selection expressions has generally remained unexploited in systemic grammar, but it will be apparent that the representations we employ constitute a graphic (more properly) lattice-theoretic correlate of such expressions. A detailed comparison between the systemic approach and our own is beyond the scope of this thesis, but once again we note the affinities



between the Prosodic-Systemic approach to language structure and that adopted in this thesis, and unification-based approaches in general.<sup>7</sup>

### 8.8. Implications for recognition

We have drawn attention to the close parallel between the lattice theoretic approach to phonotactic constraints which we advocate and Waterson's Firthian network approach. Waterson's model is in fact explicitly motivated by recognition considerations. She provides the following informal account of the way in which the organization of the child's developing lexicon into phonotactic classes might be utilized in recognition:

Context: the child is out with his mother. She draws attention to a dog carrying a stick by saying 'Look, doggie got stick' ... 'Stick' is the intonation nucleus and has the strongest accent and is thus the most salient, and is also salient by recency. The child has the word 'stick' as [gɪk]... He pays attention to the most salient part of the utterance, which is 'stick', but as [s] of [stɪk] is not auditorily salient, he identifies the phonetic pattern as plosive-vowel-plosive. He probably identifies this pattern by the sudden increase in acoustic energy on the release of the articulatory closure of the first plosive; the peak of intensity during the production of the syllabic element, which indicates a vowel, and the sudden cutting off of the syllabic phase which indicates a final plosive. The position of F1 shows a close vowel. This restricted information enables him to make a match with PVP and then a further match with P1P (see figure). No further matching is necessary: the contrast of p and k is redundant as the context, the presence of the stick, provides enough information to identify what the child has heard as [gɪk] (his 'stick') and not [bɪp] (his 'bib')... In the case of PɛP words: [bʌp], [dɒp], [gɔk], [gʌk], and [gek], the hearer's knowledge that there is a wide range of possibilities no doubt ensures that more acoustic information is included in the scanning.

Work by Zue and his associates has confirmed the utility of the classification of words based on templates employing broad phonetic classes for the purposes of recognition. In a set of studies, Shipman and Zue (1982) quantified the magnitude of the predictive power of such a classification (the results are cited in Huttenlocher and Zue 1983). Their studies examined the

---

<sup>7</sup>The interested reader may refer to Henrici (1981) for a particularly clear and explicit account of the syntax of systemic selection expressions.

phonotactic constraints of American English from the phonetic distributions in the 20,000-word Merriam Webster's Pocket Dictionary. In one study the phones of each word were mapped into one of six broad phonetic categories: vowels, stops, nasals, liquids and glides, strong fricatives, and weak fricatives. Thus, for example, the word 'speak' was represented as the pattern:

[strong fricative] [stop] [vowel] [stop]

analogous to Waterson's [plosive] [vowel] [plosive] classification. Shipman and Zue found that approximately 1/3 of the words in the 20,000-word lexicon can be uniquely specified simply in terms of this broad phonetic template. In the worst case, the template reduced the number of possible word candidates to about 1% of the lexicon.

As Shipman and Zue point out, one can view the broad phonetic classification they employ as partitioning the lexicon into equivalence classes of words sharing the same phonetic class pattern. For example, given their phonetic categories, the words 'speak' and 'steep' are in the same equivalence class. It is precisely such equivalence classes that are directly encoded in Waterson's network and in our phonotactic lattice. The lattice has a number of advantages over both Waterson's approach and Shipman and Zue's encoding, however. First, the lattice represents simultaneously a variety of classification schemes. It is possible to replicate Shipman and Zue's classification by projecting the sublattice matching their six phonetic categories. The categories they have chosen are based on manner characteristics, for the reason that 'manner differences tend to have more robust and speaker invariant acoustic cues than place differences' (p. 172). But the essence of lattice structure is cross-classification, and cross-classifying the manner-classification is a place-classification, according to which 'speak' would be encoded:

[coronal] [labial] [dorsal] [dorsal]

and 'steep' encoded

[coronal] [coronal] [dorsal] [labial]

The intersection of these classifications would uniquely identify 'speak' and 'steep'. It is of course an open and interesting question which classification schemes most efficiently categorize the set of words, and indeed what degree of resolution is required in one classification scheme, given the added discriminatory power of another — how fine-grained does the vowel

classification have to be, for example. In any case, we take the position that what is acoustically salient for recognition is a function not only the set of possibilities in the phonological system, but the acoustic nature of the background noise during recognition. In certain conditions, for example, cues to certain manner characteristics may be compromised — the distinction between strong and weak fricatives may collapse, for example — and the recognizer may compensate by ‘tuning in’ to cues for place characteristics in order to disambiguate the utterance. It is this capability that the lattice encoding facilitates.

The second major advantage of the lattice encoding, of course, is the direct representation of redundancy at all levels of resolution. To an extent this is implicit in the Zue and Shipman encoding. The fact that such a coarse classification as [strong.fricative] is sufficient to recognize the initial in ‘speak’ and ‘steep’ is of course due to the phonotactic constraint that /s/ is the only segment possible in this position.

### 8.9. Relativized redundancy

A consequence of structured specification approach to reciprocal dependency is that the phonology is agnostic with respect to which of the mutually dependent properties should be regarded as predictable, and which unpredictable. We might characterize this view as RELATIVIZED REDUNDANCY: that which is predictable is a function of what is given, or ‘known’. From a perceptual point of view, what is known of a morpheme presented for recognition is in part a function of the acoustic environment, where the signal is constantly and variably degraded by noise. In a model such as ours, in which all possible implication pathways are represented, the system of phonological redundancy can be exploited as an aid to recognition in a maximally efficient way. In the standard model, on the other hand, which is compelled to make arbitrary decisions as to what is inferable and what is not, this capacity is compromised.

Recent research in psychoacoustics also indicates that ‘what counts’ in recognition is not fixed, but varies even as a function of the *amount of attention* that is given to speech perception (Gordon et al. 1993). Specifically, cues to perception generally regarded as ‘weaker’ were found to play a greater role in perception under conditions of distraction — or of what we might term ‘cognitive noise’. In one experiment, two acoustic cues to the distinction

between /i/ (as in 'beat') and /ɪ/ (as in 'bit') were examined: formant pattern (regarded as the stronger) and duration (taken as the weaker). It was found that the relative salience of duration was of greater import under conditions of distraction. Gordon et al. conclude that speech perception is more dependent on multiple cues than was previously believed, and that it is unlikely that a single strong cue is generally dominant in recognition. A single strong cue can lead to high levels of recognition, but only if careful attention is given to the stimulus. They also note that synthetic speech which successfully encoded only the strong acoustic cues to segment identity would place heavy demands on attentional resources in order to be recognized successfully — consistent with findings regarding the difficulties listeners have in understanding synthetic speech (Luce et al. 1983).

The relevance of a phonological level of redundancy to such issues has hardly been explored — rather, attention has been focussed on the relative contributions of the phonetic as opposed to auditory levels of processing. This is in fact the point: the orthodox model of redundancy in generative phonology is ill-equipped to make any significant contribution to such research.

The position adopted in this thesis opens up a line of research focussed not simply on the role of noise in the recognition problem encountered in current speech technology, but a more 'core-linguistic' investigation of the role of noise in shaping phonological organization. We conjecture that the need to combat noise affects not just short-term speech recognition strategies, but the long-term evolution of language itself. The following paragraphs are designed to give a sense of the possibilities and scope of such a line of research.

We have in mind Shephard's notion (1984, 1987a, 1987b) that certain aspects of the world, having been invariant throughout evolutionary history, have become hard-wired into the human perceptual system: or in this case, wired into the system (speech) that interfaces with a perceptual system (hearing). Shephard (1989: 111ff) summarizes research on COLOUR CONSTANCY that illustrates the idea. The research is part of the study of a general and fundamental problem faced by any adaptive system, the problem of representing an external object as the same even though the information available about that object at the 'sensory surface' varies widely from one



presentation to the next —the problem of speech recognition can be seen as a specific instance of this fundamental problem.

Visual information available at the sensory surface about an external object varies not only with its position in space but also with the composition of the light falling on it. Mechanisms of colour constancy enable us to determine the intrinsic 'spectral reflectance properties' of the object's surface despite variations in the composition of the light that falls on the object. In other words, just as we can recognize an object as having the same shape even though it projects different shapes on the retina (as it moves), so we can recognize an object as having the same colour even though it projects different colours on the retina (as it is variously illuminated).

In terms of logical possibilities, the complete physical specification of the spectral reflectance of an arbitrary surface requires an unlimited number of parameters, giving the proportion of incident energy that is scattered back from and absorbed by the surface for each wavelength within the visible spectrum. However, Shephard (1989:111) points out that the nature of light as it appears on earth provides background constraints that severely limit this specification:

Much as a rigid shape, being constrained to an invariant three-dimensional space, has only six degrees of freedom of position (three of location and three of orientation), the spectral power distribution of natural illumination, deriving from the invariant sun, has only a few degrees of freedom of spectral composition

The spectral distribution of natural daylight has been measured over a wide range of times of day and atmospheric conditions, and has been found to be closely approximated by the linear combination of about three spectral basis functions (Maloney and Wandell 1986). Thus, spectral composition can be characterized by about just three 'degrees of freedom'; one to accommodate variations in overall level of illumination; and one or two more degrees of freedom to accommodate variations in the spectral balance between the longer wavelengths (more subject to absorption by water vapour, but more readily able to penetrate dust and other suspended particles) and shorter wavelengths (more subject to scattering by the even smaller molecules of the air itself). Shephard argues that for any adaptive system to attain colour constancy, it must analyze the input into a few colour channels — a long, medium and short-wavelength channel, say - and points out that human vision has just three classes of colour receptors, namely, cones sensitive to



longer, shorter and intermediate wavelengths. Shephard (1989:112–3) concludes:

The linear model proposed by Maloney and Wandell suggests to me that this three-dimensionality may be neither an arbitrary design feature of the human visual system nor a consequence of the surface spectral reflectances of the particular objects that were significant at various stages of our ancestral line. Three-dimensional color representation may have been favoured because three degrees of freedom are needed to compensate for natural variations in terrestrial lighting, and, thus, to achieve color constancy.

Shephard's thesis is that human perceptive mechanisms may be tuned by background constraints — such as the nature of terrestrial light — which have been constant over evolutionary history; and that this tuning, or internalizing of constraints, is integral to the utilization of the constancy mechanisms which facilitate recognition. In a similar vein, we conjecture that one of the background constraints relevant to auditory discrimination and identification is the nature of 'terrestrial noise'. The spectral distribution of the signal to noise ratio in a wide range of natural environments has only recently been systematically measured (Teder 1990). But note that constraints governing terrestrial transmission of sound have been similarly invariant over evolutionary history. In addition, auditory analogues of 'shape constancy' must certainly be relevant to the recognition of the same sound under the degradations and distortions of signal introduced by variations in distance and orientation (Wardrip-Fruin 1985).

The phenomenon of speech has this extra dimension, however, that we are not simply perceiving objects, but signs: objects whose 'purpose is perception', as it were. Accordingly, it would not be surprising if the properties required to ensure perceptual constancy were projected into the system itself. In conclusion, we conjecture that the system of phonological redundancy which pervades natural language is responsive to the need to maintain perceptual constancy in the face of variations in terrestrial noise. We propose the systematic investigation of human sound patterns in this light as a topic for future research.<sup>8</sup>

---

<sup>8</sup>It is relevant to note that Shephard's (1989) discussion of internalized constraints takes place in response to the difficulty of specifying initial structure for connectionist systems. Our discussion above thus also has implications for the connectionist modelling of speech.

In conclusion, the declarative model of redundancy structure we propose is, in a principled way, neutral with respect to the processing demands which are made upon it. The pattern of redundancy is not distorted by arbitrary decisions regarding the direction of inference between one feature and another, and the phonological representation does not give arbitrary preference to any particular subset of 'predictor' features, intuitively related to the 'strongest' or 'most salient' phonetic cues. Just as Gordon et al. propose a model in which *all* relevant acoustic cues contribute to phonetic perception, so we propose a model in which all relevant patterns of phonological redundancy can contribute to recognition. Contact can thus be reestablished between the 'deepest' levels of phonological representation and the representational models employed in speech research. Further, once such artificial distortions have been stripped away, the particular contribution of a specifically phonological level of redundancy can begin to be investigated.

## 9. Structured specification and underspecification theory

---

### 9.1. Introduction

Recent work in underspecification theory has established the central and conditioning role of redundancy and markedness in the operation of phonological rules proper, demonstrating the relevance of these notions to issues beyond the domain of phonological inventories and phonotactic constraints. However, in underspecification theory the interaction between phonological rules and statements of redundancy is formalized in terms of derivational history and rule-ordering. Predictable properties are first eliminated from representations and then reintroduced by rule. The conditioning effect is achieved by virtue of the different degree of specification that is available at the various stages of the derivation at which a particular rule applies. Certain rules will have access to representations which are **UNDERSPECIFIED** for particular features, and rule application is defined in such a way that this underspecification effects the outcome. In such a model, the same rule may apply at a later stage of the derivation to the same representation, but with redundant features inserted, and give a different result.

However, we argue that the decision to encode the redundant status of features by omission (contrastive specification), and the additional decision to encode the unmarked status of features by omission (radical underspecification) leads to intractable formal problems. In this chapter, we show how the alternative theory of structured specification allows us to express the conditioning effects of redundancy and markedness while avoiding the problems encountered in the standard theory.

### 9.2. Contrastive specification

**CONTRASTIVE** specification — also referred to as **DISTINCTIVE** specification — is defined in the following way by Archangeli (1988):

Contrastive Specification assigns specific values to a feature in underlying representation only where that feature is being used to distinguish segments in the respective contexts; non-

contrastive values are left blank. This view is advocated in Clements (1987) as well as Halle (1959) and Steriade (1987)

And Clements and Sonaiya (1989:1) contains the following characterization:

other theories [of underspecification] have proposed [...] that all distinctive feature values are underlyingly present (Steriade 1987) [...] Let us call these theories *Distinctive Feature Representation*.

Compare also Paradis and Prunet (1991:7):

In contrastive specification (e.g., see Steriade 1987; Clements 1987), the content of URs is determined by examining contrasts in the phonemic inventory of the language. If two or more segments contrast with respect to a given feature, then the feature is said to be contrastive for that pair of segments. Both members of a contrasting pair are specified for the contrastive value but, if there is no contrast, neither value is present underlyingly.

Steriade (1987) is thus repeatedly cited as a primary source for this position; and indeed that paper is notable for including an attempt at precise definitions of its central ideas. It is important to point out, however, that Steriade (1987) does *not* contain a proposal for a theory of contrastive specification; what Steriade proposes is that redundant values be omitted from underlying representations. As we have seen, however, in our examination of Halle (1959), systematically omitting redundant values results in representations which are non-distinct. Halle was explicit regarding this point, and introduced the entire machinery of branching diagrams precisely to convert non-redundant, minimally specified representations into distinct representations: an attempt which we have criticized already. The literature on 'contrastive' specification thus presents a confusing aspect: the following exposition attempts to clarify the situation.

Steriade (1987) presents a theory in which redundant (predictable non-alternating) values are omitted from representations — a position which, mindful of Halle's Condition (5), we dub minimal specification. Steriade also provides definitions which enable one to determine precisely which values should be omitted, and these definitions produce minimally specified matrices, in Halle's sense. We quote Steriade's (1987:341-2) definitions in full:

- R-class of segments with respect to F: the class of segments where a feature co-occurrence constraint blocks one value of F.
- R-value for F: the value of F present within its R-class.
- R-rule for F: a redundancy rule introducing an R-value

Let us consider a concrete example of the application of these definitions.  
 Recall the fully specified feature matrix used as an example by Halle (1959):

	t	s	c	n
nasal	-	-	-	+
strident	-	+	+	-
continuant	-	+	-	-

The feature co-occurrence constraints which obtain are as follows:

If [-strident]	then [-continuant]
If [+continuant]	then [+strident]
If [+strident]	then [-nasal]
If [+nasal]	then [-strident]

According to the definitions, each feature co-occurrence constraint determines an R-CLASS. On reflection it will be seen that the R-class is specified by the 'left-hand side' of the constraint. Thus, for example, the constraint 'If [-strident] then [-continuant]' determines an R-class with respect to [continuant], a class where one value — the positive value — is blocked. This class is just the class of [-strident] segments, {t, n}. Within this class, then, only one value of continuant is *present*, [-continuant], and this is thus the R-VALUE for [continuant]. Thus this feature value, being an R-value (read 'redundant' value) is omitted from phonological representations. These definitions may be applied with respect each and every feature co-occurrence constraint in order to determine the entire set of R-values. The resulting matrix is given below:

	t	s	c	n
nasal	-			+
strident	-		+	
continuant		+	-	

However, in the resulting matrix the pairs {t, s}, {s, n}, {c, n} — half of all the logically possible pairs — are non-distinct. It is thus apparent that Steriade's definitions are identical to Halle's definition of minimal specification. Steriade makes no reference to any machinery or conditions which would ensure distinctness, even though her definitions produce representations which are non-distinct, just as they did for Halle. To reiterate, the omission of all redundant information leads to non-distinct representations.



The declarative–procedural distinction may be pertinent here. Note that there is a procedural interpretation of Steriade’s definitions, which *may* result in distinct representations — albeit at the cost of failing to remove all redundant information. The declarative interpretation which we have taken — and taken advisedly, in the light of Steriade’s decision to make the basis of her approach feature co-occurrence *constraints* — is reminiscent of the traditional notion of ‘simultaneous rule-application’: all redundant values are identified, and simultaneously removed from the matrix. If, on the other hand, we were to interpret the feature co-occurrence constraints as a set of instructions to remove information from the matrix in a derivational manner, then we would find that in certain cases we would be *unable* to remove predictable specifications. When we come to apply a certain rule, the feature which matches the rule’s left hand side may have already been ‘erased’. Which rule fails, of course, and thus which predictable specification remains, depends on the order we choose to apply the rules, leading to indeterminacy as to the appropriate underlying matrix.

Note, however, that this procedural interpretation is not sufficient to *guarantee* distinctness in any case. Thus for the set of constraints above, interpreted as rules applying in the order given, the second rule results in the removal of [+strident] from the /s/ column; but now the third rule will not apply to /s/, since that rule’s target specification has already been removed<sup>1</sup>. The result of applying the constraints as rules in the order given is as follows:

	t	s	c	n
nasal	–	–		+
strident	–		+	
continuant		+	–	

Even on this procedural interpretation, {c, n} {t, s} are non-distinct; and we are forced to *retain* the predictable information that /s/ is [–nasal] (all stridents are non-nasal). Taking another order at random, we have:

- 1. [–strident]           →   [–continuant]
- 2. [+strident]         →   [–nasal]

---

<sup>1</sup>We thus assume a sub-matrix interpretation of rule application. See Stanley (1967) for an extensive discussion showing that a rule-based approach to redundancy is problematic on *any* interpretation of rule-application.

3. [+continuant] → [+strident]  
 4. [+nasal] → [–strident]

	t	s	c	n
nasal	–			+
strident	–		+	
continuant		+	–	

Which is equivalent to the declarative interpretation. Thus we see that procedurality alone is not sufficient to ensure distinctiveness

It might be suggested that adopting a *particular* extrinsic ordering of the constraints procedurally interpreted will ensure that the matrices remain distinct. We have not taken the time to investigate whether some particular order achieves this goal, for the following reasons. For  $n$  rules, there are  $n!$  possible orders. Thus just 10 rules would give rise to over 3.5 million rule orderings to examine. In the absence of a principled way to deduce such an ordering, such a trial and error approach seems unmotivated; and of course, even if some ordering could be found to ensure distinctiveness for a particular set of constraints, this does not ensure that one can be found in the general case. In addition, such an extrinsically ordered, rule-based approach to redundancy statements has been extensively criticized by Stanley (1967).

It is unclear whether Steriade was aware of the problem of non-distinctness that results from the systematic application of her definitions; but it does appear that she assumes that minimal specification and distinctive specification are one and the same. That is, despite the precise definitions she advances, and despite the following explicit description of her position (Steriade 1987:339):

I will suggest that only one type of predictable value is systematically absent from underlying representations: those predictable from feature co-occurrence restrictions

Steriade (1987:358) nevertheless concludes:

We have seen here that the linguistically significant boundary is that separating distinctive and non-distinctive assignments of feature values.

And indeed, subsequent researchers (such as those cited above) have taken the view that Steriade (1987) in fact advocates a theory of *distinctive* specification. Subsequently Archangeli (1988) attempted to provide a formal

algorithm for *distinctive* specification based on examination of pairwise distinctions in the segment inventory, correctly noting that no advocate of current contrastive underspecification has supplied such an algorithm. And for Paradis and Prunet, Archangeli's algorithm has become synonymous with Steriade's proposal.

The widespread belief that Steriade (1987) amounts to a theory of distinctive or contrastive specification is symptomatic of a more general assumption that pervades generative phonology: the assumption that the terms 'distinctive' and 'unpredictable' (and conversely the terms 'non-distinctive' and 'predictable') are synonymous, and mutually definable. We have seen a formulation of this position in the seminal work of Halle (1959):

an important role [...] will be played by those features and feature complexes which serve to distinguish one morpheme from another. Features and feature complexes which fulfill this function will be called *phonemic*; features and feature complexes which are distributed in accordance with a general rule of the language and hence cannot serve to distinguish one morpheme from another, will be called *non-phonemic*. [...]

Certain features are non-phonemic because they can be predicted from certain other features in the same segment

Thus 'predictable' equals 'rule-governed', equals 'non-phonemic' equals 'non-distinctive'. Recall, too, the quote from Anderson (1985:10) with which we began our investigation of redundancy in Chapter 4:

It is often taken as self-evident in phonological studies that underlying representations should contain only *distinctive* or nonredundant material

The equation is succinctly formulated by Kenstowicz and Kisseberth (1977:131-2):

In any given language only certain features of pronunciation are distinctive — that is, unpredictable and thus capable of distinguishing between lexical items in underlying representations. The remaining features are non-distinctive — that is, predictable and thus incapable of distinguishing one underlying form from another

This equation is false. Indeed, as we have seen, the very framework of Halle (1959) recognized as much, where it was found that certain predictable information had to be supplied *by rule* precisely in order to satisfy the distinctness condition. Adherence to this equation is so entrenched, however, that its rebuttal deserves further space.

Consider once again the fully specified matrix:

	t	s	c	n
nasal	–	–	–	+
strident	–	+	+	–
continuant	–	+	–	–

We ask the following question: What *distinguishes* /c/ from /n/? The answer is clear : [nasal] or [strident] (or both). Note particularly the disjunction (or conjunction) in the answer: /c/ and /n/ are distinct with respect to *both* [nasal] and [strident]. In Halle’s terminology, we can say that the FEATURE COMPLEX [nasal, strident] distinguishes /c/ and /n/.

We now ask the following question:

- (i) What is *predictable* for /c/ and
- (ii) What is *predictable* for /n/?

The answer is again clear:

- (i) For /c/, [–nas] is predictable (all stridents are oral) and
- (ii) for /n/, [–strident] and [–continuant] are predictable (since /n/ is the only nasal, all other features are predictable).

We thus have the following results:

	<i>Distinctive</i>	<i>Predictable</i>
/c/	nasal, strident	nasal
/n/	nasal, strident	strident, continuant

We see immediately that, for each segment, one and the same feature is *both* distinctive and predictable. Thus, as we have said many times, if one attempts to omit all predictable information, one must necessarily omit certain distinctive information, and representations may become non-distinct.

Nor, we hasten to add, is this a quirk of the particular example which we have been using for illustration. Consider the canonical five-vowel system, classified by the features [high], [low] and [back]:

	i	e	a	o	u
high	+	–	–	–	+
low	–	–	+	–	–
back	–	–	+	+	+



We may ask similar questions: What distinguishes /a/ and /u/? Clearly, the features [high] and [low]. What is predictable for /a/? The values for [high] and [back]. What is predictable for /u/? The value for [low]. We thus have the following results:

	<i>Distinctive</i>	<i>Predictable</i>
/a/	high, low	high
/u/	high, low	low

Thus for /a/, [high] is both predictable *and* distinctive; for /u/, [low] is both predictable *and* distinctive.

Interestingly, Archangeli shows that application of the particular algorithm she proposes on behalf of advocates of distinctive specification fails to achieve the desired goal: it leads in certain cases to non-distinct representations. She then speculates on ways to modify the algorithm to meet this problem, observing that any such modifications will lead to indeterminacy in the representation. Archangeli (1988:202) concludes ‘These are serious questions, and issues that remain to be dealt with in Contrastive Specification Theory.’ As we have seen, however, precisely such questions had already been addressed in the distinctive specification theory of Halle; and the theory of branching diagrams consitutes an algorithm of just the sort Archangeli seeks. However, as we have also seen, it suffers from the problem of indeterminacy.

We see then that the position adopted as ‘contrastive specification’ in current phonology does not improve in any way over the position advocated by Halle (1959); it is, in fact, significantly less theoretically developed, failing even to distinguish MINIMAL from DISTINCTIVE specification.

However, putting to one side these theoretical issues for the moment, Steriade (1987) is important for another reason. The greater part of her paper is a collection of empirical data which conclusively demonstrate that phonological rules are sensitive to whether the features they manipulate are redundant or non-redundant. We provide one such example below (Steriade 1987:343).

In the Pasiego dialect of Montañes Spanish a harmony rule assimilates non-low vowels to the height of a non-low stressed vowel. However, low vowels do not undergo, block or trigger the rule. Examples are given below:



/beb/ drink  
bib-ís  
beb-émus  
beb-ámus

/sint/ feel  
sint-ís  
sent-émus  
sint-áis

The underlying form of the stem vowel surfaces only when followed by a [+low] vowel. Steriade argues that the neutrality of /a/ in Montañés Spanish is correlated with the fact that the [-high] specification is redundant just in the case of [+low] vowels:

The lack of trigger or blocker behaviour of low vowels in the height harmony of Pasiego indicates that these vowels lack height values at the stage when harmony applies. That low vowels cannot be underlyingly associated with any value for [high] follows from the fact that they represent the R-class of segments for [high]

We agree that Steriade successfully establishes the fact the behaviour of phonological features is conditioned by their redundant/non-redundant status; we have also seen however that the notion of contrastive specification is inherently flawed. One possible response would, of course, be to replace the notion of contrastive specification with that of minimal specification, and to admit non-distinct representations in underlying forms. This would in fact bring the theory a step closer to radical underspecification, which we examine in the following section, and which does admit non-distinct representations. This step has not, to our knowledge, been suggested.<sup>2</sup> However, as we will see, quite independently of the issues which concern underspecification theory, there are a number of arguments *against* the use of blanks in phonological representations — against, that is, the decision to encode redundancy by omission. In fact, we have already presented one such argument:<sup>3</sup> omitting redundant specifications leads to geometrically discontinuous representations, under current interpretations of feature geometry. Moreover, as we will see, the decision to use the apparatus of

---

<sup>2</sup>Stanley (1967:421) suggests an alternative to the distinctness condition, the 'true-generalization condition', which is immune from criticism. However, crucial to its formulation is the assumption that redundancy rules be disjoint from phonological rules, in the sense that all the former apply before any of the latter. In a model such as the envisaged minimal underspecification, in which underspecified matrices are subject to the full-power of phonological rules, this solution would not be tenable.

<sup>3</sup>Chapter 7, §7.3.

blanks and insertion rules leads to internal problems and incorrect predictions in underspecification theory itself.

Current work in contrastive underspecification has simply fallen back on the only technique for coding redundancy into representations that has ever been suggested in generative phonology — underspecification. However, simply because it is possible to condition phonological rules to the redundant vs non-redundant status of the features they manipulate using the mechanism of blanks and blank-filling redundancy rules, this does not mean that it is necessary to use this machinery. All that is required is an alternative, principled way of encoding the relevant distinctions. What is required is a 'blank-free' approach to redundancy, which nevertheless codes redundancy relations in a principled way. This is what the theory of structured specification provides.

### **9.3. Radical underspecification**

Steriade (1987) is in fact a response to an alternative theory of underspecification: RADICAL UNDERSPECIFICATION (Kiparsky 1982, Pulleyblank 1983, Archangeli 1984; see also Archangeli 1988 for an overview). As the name implies, this approach omits even more specifications than distinctive specification; indeed, omits even more than minimal specification.

The major difference, for our purposes, is that in addition to omitting specifications which are predictable due to feature co-occurrence constraints (for which we reserve the term *redundant features*), radical underspecification also omits 'unmarked' or 'default' feature specifications. For any binary feature, one value — the unmarked or default value — is omitted from representations. This value cannot be inferred from any other feature in the representation — there is no implicational relationship to be expressed — it is simply a question of indicating in the phonology which value is to be taken as (literally) marked, and which value is to be (literally) unmarked. The charge that such an approach is overly stipulative is met by assuming that default values are (generally) determined by Universal Grammar. Mohanan (1991) identifies a number of variations on the theme of radical underspecification which have been suggested in the literature. For our purposes, these variations among these theories are not as significant as their common core: the decision to encode the default status of features by underspecification.

An example of a radically underspecified matrix, and the associated rules, are given below (Archangeli 1988:193):

	i	e	a	o	u
high		-		-	
low			+		
back				+	+

- [+low] → [-high]
- [+low] → [+back]
- [] → [-low]
- [] → [+high]
- [] → [-back]

Thus, in addition to redundancy rules of the familiar sort, we also have three rules — one for each feature — which indicate that the unmarked, default values for [low], [high] and [back] are [-low], [+high] and [-back] respectively.

We may summarize our position with respect to this theory in the following way. We argue that, just as redundant values should not be omitted from underlying representations, but explicitly coded as redundant through the use of structured specification, so default values should not be omitted from underlying representations, but explicitly coded as default values — ultimately by incorporating defaults into the hierarchy of features, as suggested in the previous chapter. As we have seen, however, this is not a trivial task; but then we feel that the very difficulties that have been encountered in this area by researchers in knowledge representation show that there are significant and unresolved problems concerning the interaction of implication and defaults which have been largely ignored by phonologists.

However, just as we employ the logic of implication to ground the structured specification approach, so we will employ ideas from the emerging field of DEFAULT LOGIC to explicate the issues surrounding phonological defaults. We will discuss specific examples of the use of this logic — if not its hierarchical interpretation — in the subsequent chapter.

#### 9.4. Structured specification vs contrastive specification

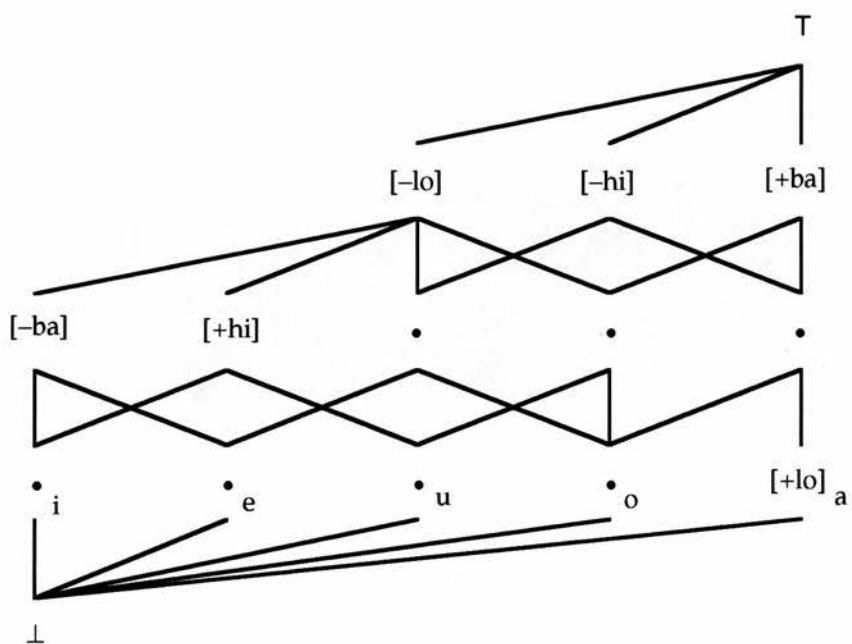
In subsequent sections, we will raise a number of objections to the underspecification approach to redundancy and defaults, and show how the theory of structured specification, and the logic of defaults, avoids these problems. Before we proceed to these comparisons, we first indicate how the theory of structured specification makes the distinctions necessary to handle the kind of empirical data adduced by Steriade.

We first note a similarity between Steriade's approach (at least, the approach as set forward in her definitions) and our own. Recall Steriade's explicit claim that *predictable* specifications should be omitted from underlying representations (not *non-distinctive* specifications), and bases her definitions on the system's feature co-occurrence constraints. Our approach, too, takes as primary the predictable/non-predictable classification of features: recall that the structural relations of the lattice may be interpreted as a direct encoding of the logical implications that hold among the various features. Secondly, note that Steriade's account of Montañes Spanish relies on being able to exclude, in a principled manner, a particular segment from the class of segments referred to by the rule: the segment for which the value of [high] is predictable, /a/. Accordingly, we will be able to express the leading idea of Steriade's analysis if our formalism allows us to omit /a/ from the class of segments mentioned in the rule, and to omit it *by virtue of* the predictable status of [high] for that segment.

The relevant lattice for Montañes Spanish is as follows<sup>4</sup>:

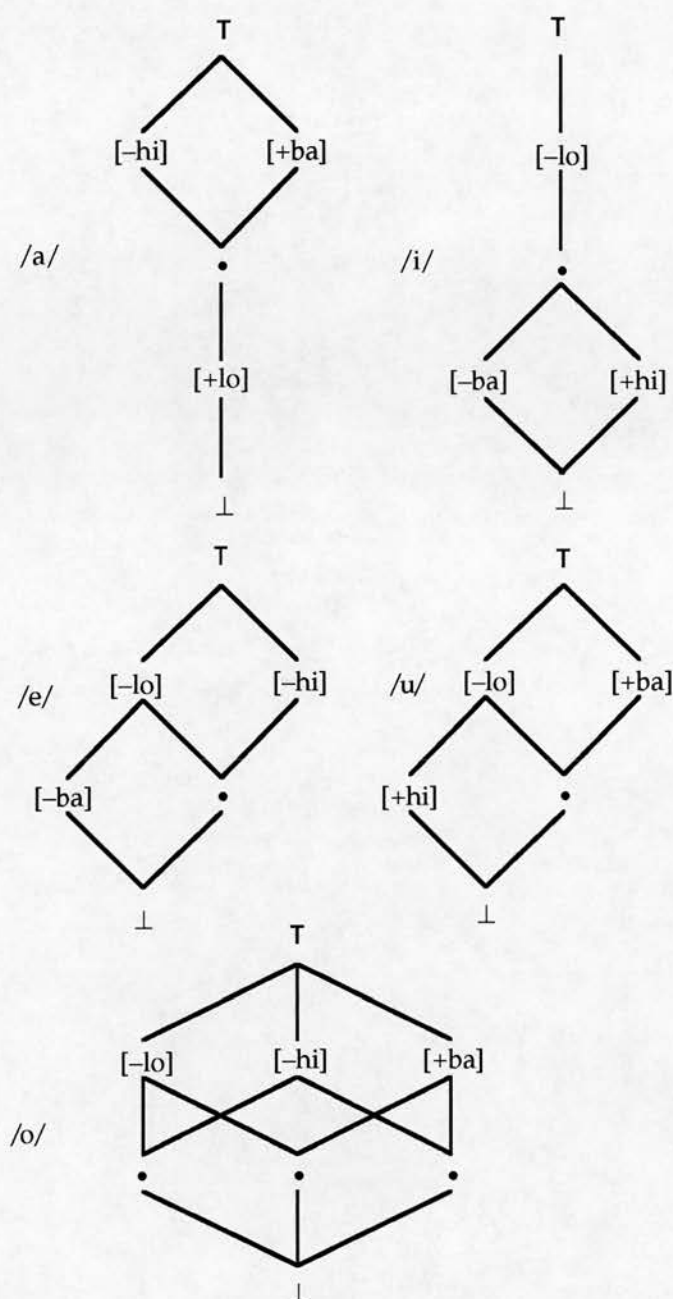
---

<sup>4</sup>In fact, Montañes is not a five-vowel system, as the lattice indicates, but a nine-vowel system: a five vowel system in which all but one vowel comes in two forms, variously analyzed as [±tense], [±ATR], [±fronted]. We have suppressed this dimension for expository purposes. The extra feature does not effect the relations between [high] and [low] which are crucial to the argument.



The representations which this lattice induces are as follows:





As can be seen, the non-predictable features are simply the lowest features in each of the lattices. As can also be verified,  $[\pm\text{back}]$  is non-predictable for every segment except /a/. The approach provides, as it were, a structural characterization of redundancy, and phonological rules can formally be sensitized to redundancy by mentioning this specification structure in their structural descriptions. We feel that little purpose would be served in developing a graphical phonological rule notation to manipulate these structures; recall that the diagrams above are not lattices, but *diagrams* of lattices, and can be drawn in a variety of ways: we do not wish to place

undue focus on properties of the *pictures* of phonological representations. The crucial point is that the intuitive notion of 'non-predictable feature' can be formally expressed as 'least meet-irreducible element', and that phonological generalizations can require that 'spreading' or 'blocking' elements bear this structural relation.

#### 9.4.1. *System geometry*

The representations above reveal a striking and surprising relationships between the five vowels. For example, /i/ and /a/ are structural inverses of each other, as are /e/ and /u/. /o/ stands apart, and is in some sense most complex. Recall, too, that these representations are a function of the constraints that characterize the phonological system of the language: segment structure is a function of system structure. This suggests an entirely new taxonomy of segment types, cutting across existing schemes of classification, be they feature based, or 'particle' based; and may make a contribution to such questions as, What makes a vowel system natural? To what extent do systems favour binary contrasts? Why do {i, a, u} form a natural distributional class? What is the basis of certain observed vowel asymmetries, categorical and statistical? What are the parameters of synchronic and diachronic vowel merger?<sup>5</sup> The development of these ideas awaits further study.

### 9.5. Problems with underspecification

We have made repeated reference in this and preceding chapters to the problems that follow from the decision to encode redundancy by omission. In this sections we illustrate these problems; some are theory internal to the constraint-based approach to phonology, others hold of any phonological framework. However, the first such argument we consider is not, we claim, a valid objection at all.

#### 9.5.1. *The Stanley-Lightner argument and ternary-power*

The most celebrated argument for the rejection of blanks to encode redundant information is the so-called Stanley-Lightner argument, and indeed it is

---

<sup>5</sup>See, Goldsmith (1987) for an investigation of a number of intriguing questions concerning segment systems.

repeatedly cited in the current underspecification literature (see, for example, Archangeli 1988). As we have just indicated, even though this argument is admitted by many researchers in underspecification to be a compelling argument against the use of blanks, we do not treat it as such. We referred to this issue in passing in the introductory chapter. Stanley argues that the unguarded use of blanks effectively permits us to refer to the set of sounds which have '0' for some feature F as opposed to the set of sounds which have '+' for that feature (in exactly the same way that the theory of binary features allows us to refer to the set of sounds which have '-' for some feature F as opposed to the set of sounds which have '+' for that feature). As Stanley (1967:410) puts it:

in using this system we must take care that the feature values remain BINARY, and that '0' in a dictionary matrix is never allowed to function as a third value, distinct from '+' and '-'. The correctness of any empirical claim that distinctive features are binary is, of course, not at issue here. The point is simply that, once we decide to use a binary system, we must be formally consistent.

In response to this argument, commentators agree that the 'undesirable consequence' that results from allowing rules to apply to underspecified matrices is that a 'ternary distinction' is permitted. Many underspecification accounts (e.g. Kiparsky 1982, Pulleyblank 1983, Archangeli 1984, 1988) have taken the view that Stanley's arguments provide compelling evidence against the use of ternary oppositions, and have thus attempted to argue that although they employ blanks in their representations, independent properties of their frameworks ensure that the use of blanks does not lead to ternary power. Thus Archangeli and Pulleyblank (1989) argue that at any *one point* in the derivation, a feature contrasts either with a blank (prior to specification) or to its opposite (after specification). Thus, at any point in the derivation, the distinction is always binary: there is never any point at which plus, minus and 'zero' are all present.

Note, however, that Stanley does not simply present an argument against blanks encoding ternary distinctions. Indeed, he says explicitly that this is not the point. His argument is, that *if* we wish our phonology to be confined to binary distinctions, *then* we must avoid the use of blanks to (covertly) encode ternary distinctions. Recall, however, that for Stanley, a sound will have '0' for the specification of a feature just in case that specification is PREDICTABLE,

or REDUNDANT. Therefore, Stanley's charge is that the use of blanks permits reference to the set of sounds for which *F* is *predictable* as opposed to the set of sounds for which *F* is unpredictable. This is a class which is outside the set of natural classes which binary classification is intended to capture; and thus the use of blanks was rejected by Stanley.

Now as we have just seen, one of the central claims of current underspecification theory is precisely that there *is* empirical evidence that shows that it is necessary to interpret features as making more than a simple binary distinction. That is, rather than segments being 'F' or 'not F', underspecification theory claims that segments behave differently depending on whether they are 'F', 'not F' or '*predictably* F'. And as we have also seen, in contrastive underspecification rules are sensitized to the predictable vs non-predictable status of the features they manipulate precisely by OMITTING redundant specifications. We thus conclude that Stanley's famous argument against allowing phonological rules access to representations with blanks is no longer compelling.

If, however, there are *independent* arguments against the use of blanks to code redundancy, then the conclusion will be that, despite the initial attraction to the idea of using blanks for the coding of the three-way distinction mentioned above, it should nevertheless be resisted, and an alternative approach to specification preferred.

One such piece of evidence is that there is *another* three way distinction that must be encoded using blanks. The use of blanks we have considered so far is the use which indicates that the value of some feature [F] is predictable since fixed and predetermined for a segment. This second use, on the contrary, is intended to indicate that the value of [F] is predictable since it varies and is underdetermined for a segment (or better, class of segments): it denotes an archisegment.

#### 9.5.2. *Ambiguity I: archi-segments*

We mentioned above that some researchers have explicitly admitted the possibility of allowing three-way distinction in phonological representations. Clements is one such. However, when this position is examined, we find that — even though the distinction is expressed in the same way as the Stanley-



Lightner argument, as involving [+F], [-F] and [øF] — the meaning of this particular three-way distinction is not '[F], [not-F], [redundantly-F]'; but '[F], [not-F], [indeterminate-for-F]'. Thus Clements (1985:242):

the consonant/vowel asymmetries noted above can be explained within representational systems that allow three-way distinctions between segments characterized by [+F], [-F] and [øF] (or absence of F), for one or more features F. In the context of autosegmental analyses, we frequently find motivation for recognizing such underlying three-way distinctions, such as that between high-toned, low-toned and toneless vowels. Under theories requiring all segments to be fully characterized for all features, there is no straightforward way of representing such a three-way opposition. In autosegmental phonology, which is not subject to such a constraint, such distinctions can be easily captured on the assumption that some segments are 'incompletely characterized' by certain features — that is, not linked to any occurrence of such features on the relevant autosegmental tier<sup>6</sup>

Note the relevant three-way distinction is 'high-toned, low-toned, and toneless'; not, say, 'high-toned, low-toned, and redundantly high-toned'. That is, interpretation of the blank here is *not* the interpretation discussed by Stanley; the one which forms the focus of the Stanley-Lightner argument. As Stanley (1967:410) reminds us 'What is important is that we keep the meaning of '0' clearly in mind'. The blank referred to by Clements indicates a segment which, in underlying representation, has a *choice* of completions; the particular choice is determined by context, and is in that sense predictable. In the case of redundant blanks, however, the blank indicates a segment which, in underlying representation, has its specification for F 'already decided', as it were. In the former case, involving an underdetermined specification, the accompanying features are insufficient to determine what value it should assume: that is why it is left blank. In the case of predetermined specification, on the other hand, the accompanying features are *totally* sufficient to predict what value it will assume; and *that* is why it is left blank. The case of

---

<sup>6</sup>We note in passing that the notion of 'uncharacterized segment' is not exclusively formulable in terms of autosegmental phonology. It is simply that the theory of autosegmental phonology happens to be a theory which includes this possibility. The suggestion that segmental phonology of the SPE-type include such uncharacterized segments was made by, for example, Shibatani and Crothers (1974). And we have seen that Halle (1959) advocated the inclusion of archisegments in his model, although his particular formulation is inadequate.



underdetermination is precisely the case we have discussed earlier under the heading of archisegments. We thus have to do in phonology with two *different* 'three-way distinctions'; and in current phonological theory they are both encoded in the same way, by the use of blanks.

### 9.5.3. *Ambiguity II: default values*

Moreover, this ambiguity is compounded in the theory of radical underspecification, for in this theory blanks are also used to encode default values. Thus the distinction '[+F], [-F], [øF]' is also made to express the three-way distinction 'F, not-F, F-by-default'. Thus in addition to using blanks to encode predetermined values, this theory also employs blanks to encode underdetermined values and default values. On the contrary, we are proposing a theory which distinguishes each such variety of specification: (i) redundant specification (ii) partial specification (iii) default specification, where the unifying notion behind these three sorts of specification is subsumption, as expressed in the hierarchical structure of the lattice. As we will now see, the failure to formally distinguish these distinct types of specification leads to problems.

One of the most striking claims of radical underspecification theory is that, since every feature has a default value, the segment containing just those features will constitute a default segment, and might be expected to behave asymmetrically from other segments in the inventory. The epenthetic segment, for example, has been shown in a number of languages to correspond to this default segment.<sup>7</sup> In underspecification theory, which encodes default feature values by blanks, the default segment will be totally underspecified.

However, as we have seen, current phonological theory also encodes archisegments by the use of blanks, and the 'archi-vowel', which is underdetermined with respect to all place features, will be representationally identical to the default vowel. Theories which do not distinguish these two types of specification, then, will be unable to distinguish these two types of segment. This would be a desirable result, if it were the case that archivowels and default vowels behaved identically with respect to phonological rules,

---

<sup>7</sup> But see Mohanan (1991:318) for criticism of this correlation.

and did not need to be formally distinguished. However, we now present evidence from Klamath and Basque, both of which languages have an archivowel in addition to a default vowel, which shows that the two types of segment in fact behave differently, and need to be distinguished representationally.

•Klamath

We first note that Klamath exhibits a number of processes which insert /a/ in phonological environments. As Kisseberth (Kisseberth 1972:10) observes 'Many four-consonant clusters undergo an /a/-epenthesis rule':<sup>8</sup>

$$\check{c}'in + w + bg + a \Rightarrow \check{c}'inwabga$$

Kisseberth (1972:10) continues 'I cannot possibly consider in detail the various phonologically conditioned /a/-epenthesis processes'. It is for precisely such a phenomenon — the recurrence of the same vowel across a number of epenthesis rules — that radical underspecification accounts for by the notion of a default segment. If each distinct epenthesis rule has to mention a fully-specified vowel in its structural description, then it is merely accidental that the same vowel should occur in each rule. If, on the other hand, epenthesis rules simply insert a 'V-slot' — underspecified for any place features — then all such segments will be assigned default values late in the derivation, irrespective of their source. The default values for Klamath, then, are such as to yield the segment /a/, inserted by the following redundancy rules:

- [] → [+low]
- [] → [-high]
- [] → [-round]

As a corollary, *non*-epenthetic /a/ will be totally underspecified in underlying representation, since all its values are unmarked values.

Klamath has a number of prefixes which contain a 'variable vowel', which harmonizes with the initial vowel of the stem, e.g. **hVs-** 'direct causative', **snV-** 'indirect causative', and **sV-** 'reciprocal' (Kisseberth 1972: 4):

<i>Indicative</i>		<i>Causative 1</i>
<b>peew-a</b>	bathes	<b>hes-peew-a</b>

---

<sup>8</sup> All Klamath data is taken from Kisseberth's paper.

<b>noog-a</b>	is cooked	<b>hos-noog-a</b>
<b>m'aas?-a</b>	is sick	<b>has-m'aas?-a</b>
<i>Indicative</i>		<i>Causative 2</i>
<b>q'eegi</b>	is absent	<b>sne-q'eegi</b>
<b>qdooč-a</b>	rains	<b>sno-qdooč-a</b>
<b>tsaaktgi</b>	gets light	<b>sna-tsaaktgi</b>
<i>Indicative</i>		<i>Reciprocal</i>
<b>n'eest'-a</b>	has sexual intercourse from behind	<b>se-n'eest'-a</b>
<b>loočw'-a</b>	covets	<b>so-loočw'-a</b>
<b>twaaq'-a</b>	smears	<b>sa-twaaq'-a</b>

Kisseberth (1971:5–6), after demonstrating that this vowel must be underlying, rather than epenthetic, comments:

If the variable vowel is part of the underlying representation of the three prefixes, it is necessary to determine precisely what representation this vowel is to be given. Since the vowel has no inherent properties of its own (its quality being a function of the following vowel), no language-internal evidence exists for assigning any specific phonetic features to the basic representation of this vowel. It might be suggested that the variable vowel be entered in the lexicon simply as the unmarked vowel, i.e. /a/. But additional data considered below shows that the morphophonemics of the variable vowel are distinct from those of basic /a/ in identical contexts. Consequently, the variable vowel cannot be identified with /a/. Nor is there any other vowel in the system that provides a suitable substitute.

I will assume that the variable vowel is, in underlying representation, a kind of pro-vowel, an archiphoneme specified as syllabic, but otherwise having no inherent feature specifications [...] I will use the ad hoc symbol V\* to stand for the pro-vowel [...] I use it in order to leave open the question of how the notion of pro-vowel is to be made explicit in a theory of grammar.

The relevance of Klamath to our argument is immediately apparent. Kisseberth explicitly claims that the variable vowel — an underdetermined or archi-segment — must be representationally distinguished from the 'unmarked' vowel — a default segment — on the basis of the different behaviour they exhibit in identical contexts.

How does this difference manifest itself? The pro-vowel also occurs in an additional class of morphological elements, — classificatory morphemes. Unlike prefixes, which always attach to consonant initial stems (Klamath words are all consonant initial) classificatory morphemes may attach to either

consonant initial or vowel initial morphemes — bound forms expressing locative, aspectual and directional meanings. Classificatory prefixes have three prosodic/melodic types, which may be represented schematically as follows:

- (i) C-                      (ii) CV\*-                      (iii) Ci -

The first type consists of a bare consonant, and is exemplified by the morpheme **w** 'act with a long instrument':

<b>w-aq'aaq'-a</b>	hangs a long object around someone's neck
<b>w-elg-a</b>	knocks down with a long instrument
<b>w-oliin-a</b>	peels off with a long instrument
<b>w-p'ak'a</b>	smashes with a long instrument

The second type consists of a consonant and the variable vowel, and is exemplified by IV\* 'act with a round object'. Note that when such a morpheme combines with a vowel initial morpheme, the vowel of that morpheme surfaces intact:

<b>l-aq'aq'-a</b>	hangs a round object around someone's neck
<b>l-elg-a</b>	puts down a round object
<b>l-oliin-a</b>	takes a round object off the edge
<b>li-tt'a-</b>	split open with a round object (stem <b>-tit'-</b> )

The final type of classificatory morpheme has a unique vowel associated with it, exemplified by **ʔi** 'act upon plural objects'. Note that when such a morpheme combines with a vowel initial morpheme, the initial vowel of that morpheme is dropped:

$\text{ʔi} + \text{elga} \Rightarrow \text{ʔilga}$   
 $\text{ʔi} + \text{odiila} \Rightarrow \text{ʔidiila}$

Another such classificatory morpheme is **č'a** 'act on a handful of granular objects'. As Kisseberth (1972: 9) notes 'it is the existence of a morpheme like **č'a** that shows that the pro-vowel cannot be equated with the unmarked vowel /a/'. Compare the behaviour of **IV\*** and **č'a** in combination with the vowel initial morpheme **oy** 'give':

$$\begin{array}{lcl} \text{IV}^* + \text{oy} + \text{a} & \Rightarrow & \text{loya} \\ \check{\text{c}}'\text{a} + \text{oy} + \text{a} & \Rightarrow & \check{\text{c}}'\text{aya} \end{array}$$



Thus *č'a* behaves like a morpheme ending in a full vowel, rather than like one ending in the pro-vowel. Accordingly, the pro-vowel and the default vowel must be distinguished in underlying representation, but in the theory of radical underspecification, which encodes both partial and default specification using the same mechanism, this is impossible.

•*Basque*

Exactly the same problem for radical underspecification has recently been noted by Hualde (1991:205), with respect to Arbizu Basque:

Here I show that, contrary to the predictions of radical underspecification, there are languages where a distinction must be made between vowels with no features (empty V slots) and the unmarked vowel

Hualde first notes that in Arbizu /e/ consistently behaves as the unmarked vowel, being inserted in all instances of epenthesis. For example, when the suffix *-k* 'ergative indefinite' is suffixed a consonant final stem, /e/ is inserted to break up the cluster:

<i>Base</i>	<i>Erg. indef.</i>	
<b>alaβa</b>	<b>alaβaak</b>	daughter
<b>paate</b>	<b>paatek</b>	wall
<b>ašto</b>	<b>aštok</b>	donkey
<b>mendi</b>	<b>mendik</b>	mountain
<b>ešku</b>	<b>eškuk</b>	hand
<b>čakur</b>	<b>čakurek</b>	dog
<b>gison</b>	<b>gisonek</b>	man

Like Klamath, Arbizu Basque also exhibits a type of affix which contains a variable vowel: one such is *-V\*n* 'genitive definite'. Note that, when combined with this pro-vowel, the vowel of the stem simply spreads into the suffix:

<i>Base</i>	<i>Gen. indef.</i>
<b>alaβa</b>	<b>alaβaan</b>
<b>paate</b>	<b>paatee</b>
<b>ašto</b>	<b>aštoon</b>
<b>mendi</b>	<b>mendiin</b>
<b>ešku</b>	<b>eškuun</b>
<b>čakur</b>	<b>čkuren</b>
<b>gison</b>	<b>gisonen</b>

The pattern when a full vowelised suffix is combined with a vowel initial stem is quite different. There are a number of rules affecting such a combination: if



empirical step back from pure pro-vowels, we find vowels which harmonize for all but one feature, such as the Efik singular prefix, which adopts most of the features of the stem initial vowel, but is inherently [–high] (Cook 1986). Exactly the same ambiguity of specification arises in this scenario as happens in the totally underspecified case.

As we have seen previously, Halle (1959) attempted to provide a formal definition of the archiphoneme based on branching diagrams and the notion ‘incomplete specification’, an attempt which was (properly) rejected by Stanley (1967). In the full-specification approach advocated by Stanley and SPE, the archiphoneme had no formal status. In current phonological theory, it has again been proposed that the notion of underdetermined segment be given formal expression. But the means that have been adopted — the use of blanks in underlying representation — leads to an undesirable formal ambiguity between archi-segments and unmarked segments.

In our approach, on the other hand, the notion of the archiphoneme is expressed by partial specification, while the notion of the default segment expressed by default specification. While these notions are formally distinct, they are also shown to be formally related, by virtue of the subsumption relation. An archiphoneme is identified with a node relatively high in the redundancy hierarchy, a node which subsumes the nodes associated with its more fully instantiated completions. In a similar way, a default segment is identified with a node relatively high in the markedness hierarchy, a node which subsumes the nodes associated with more exceptional segments.

#### 9.5.4. *Ambiguity III: Monovalent features*

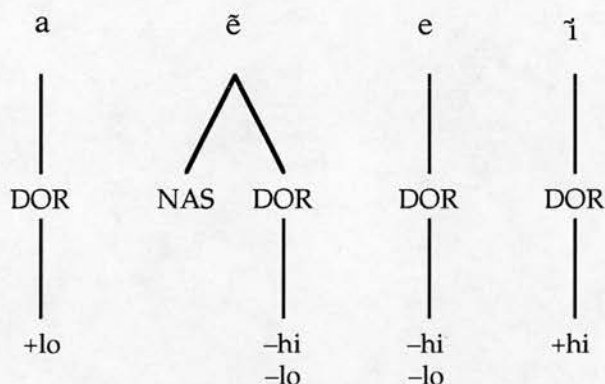
In fact, the current state of phonological theory is even more parlous than the previous discussion indicates. There is a *fourth* type of three-way distinction which is made in current phonology, and this distinction is *also* encoded using blanks. This is this three-way distinction ‘F’, ‘not F’, ‘undefined-for-F’. In current generative phonological theory, this notion has been expressed using the device of ‘monovalent’ or ‘unary’ features, and leads to logical inconsistencies when combined with the decision to simultaneously use blanks to encode redundant values.

Consider the following segment inventory:

	a	e	o	ẽ	õ	ĩ	ũ
high	-	-	-	-	-	+	+
low	+	-	-	-	-	-	-
back	-	-	+	-	+	-	+
nasal	-	-	-	+	+	+	+

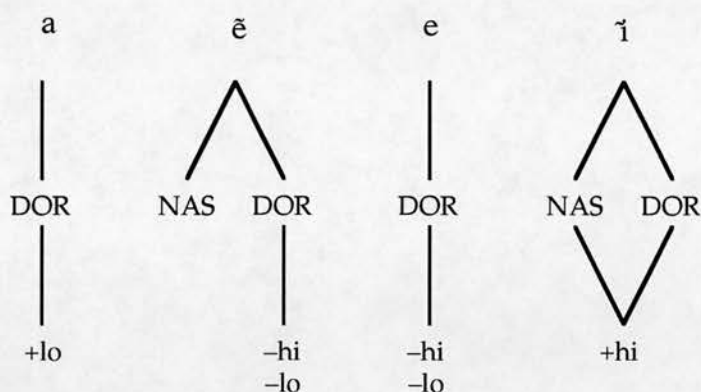
Note that nasality is distinctive for mid-vowels, but predictable for low vowels and high vowels: low vowels are oral, high vowels are nasal.

Suppose we choose to encode nasality as a monovalent feature, and also, as in the standard approach, encode redundancy by omission. Employing the standard representational apparatus of feature geometry, and omitting predictable information, the representations for /a/, /e/, /ẽ/ and /ĩ/ will be (irrelevant structure omitted):



The distinctively nasal segments are {ẽ, õ}; the distinctively oral segments are {e, o}. The former can be identified structurally under this approach, through the presence of the nasal node; but the distinctively oral segments can neither be distinguished from segments which are redundantly oral, nor from segments which are redundantly *nasal*.

Now consider the representations which result from a structured specification approach (again, irrelevant structure is suppressed):



This is a striking and, perhaps, a surprising result. Again, it turns out that it is not possible to distinguish segments which are distinctively oral from segments which are redundantly oral; but it *is* possible to distinguish segments which are distinctively oral from those that are redundantly nasal.<sup>9</sup>

This latter result is clearly desirable: no-one would wish to claim that /e/ and /ɪ/ were members of the same natural class with respect to nasality, whether nasality is considered distinctively or *simpliciter*. But this is precisely what the requirement to omit redundant information predicts. Thus, just as we have seen that there are undesirable consequences of combining underspecification and the sub-classificatory aspects of feature geometry, so we now see there are undesirable consequences of combining underspecification and the monovalent aspects of that theory. And further, we see that the problem is solved by adopting the theory of structured specification.

Moreover, having clarified the situation by removing this potential confusion we see that a striking prediction is made with respect to the interaction of distinctiveness and monovalency. That is, it is not possible formally to distinguish *distinctive* 'monovalent absence' from *redundant* 'monovalent absence': the distinctive/redundant dimension is neutralized in the case of monovalent absence. There is thus an asymmetry regarding the distinctive/redundant relation in the case of monovalent specifications: more succinctly, the presence of a monovalent specification may be distinctive or redundant; its absence can be neither. The problem with the standard theory

---

<sup>9</sup> As previously discussed, distinctively nasal segments can be distinguished from redundantly nasal segments by the structural position of NASAL in the hierarchy.

is that it conflates 'redundant presence' with (monovalent) 'absence', thus confusing *redundantly nasal* segments and *non-nasal* segments.

Thus we see that the standard theory conflates four separate dimensions of specification: predictable values, default values, potential but undefined values, undefinable values. The standard theory uses one and the same formal device — underspecification — to express all four cases of specification. The theory of structured specification encodes redundant values by inheritance; default values by default inheritance; archival values by non-terminal nodes; and privative values by non-complemented nodes.

#### 9.5.5. *The problem of reciprocal dependencies*

Another independent reason for not expressing redundancy through omission is the problem of mutual dependency. We have discussed this issue elsewhere, and will not repeat the discussion here. We simply point out this extra consequence of the reciprocal redundancy problem: it shows that redundancy cannot be expressed by underspecification; and this in turn shows that although it is possible (in certain cases) to sensitize rules to redundancy by using blanks, this possibility is nevertheless to be resisted.

There are two other reasons for rejecting the underspecification mechanism as a way of sensitizing phonological rules to the distinctive/redundant status of features. The first of these reasons is theory internal to a unification-based approach to phonology, and to any approach which shares its conception of rule application. The second has wider application.

#### 9.5.6. *Rule-application I: sub-matrix application*

As noted by Stanley (1967:416) a theory which sensitizes rules to the redundant/distinctive status of features by using blanks must crucially employ a SUBMATRIX INTERPRETATION of rule application.

The term is due to Stanley himself (Stanley 1967:413). He argues that a natural formulation of the notion of rule application is 'If a matrix M meets the structural description of a rule SD(R), then it must also meet the structural change of that rule SC(R)'. Formally, however, he shows that there are two possible interpretations of what it means to 'meet' the structural description of a rule, and thus two interpretations of rule application.



The first he terms the SUBMATRIX INTERPRETATION. Let M and N be two feature matrices; then M is a SUBMATRIX of N if, whenever M has a specification ('+' or '-') for some feature, N has the same specification, but not necessarily conversely. We can then say that M MEETS SD(R) iff SD(R) is a submatrix of M, and thus R applies to M.

The second interpretation of rule application is the DISTINCTNESS INTERPRETATION. Stanley defines matrices M and N to be DISTINCT if some feature F has opposite values in M and N (a notion first introduced by Halle (1959). However, there, the relation is defined between two representations, whereas here it is defined between the *structural description of a rule* and a (target) representation. We then say a rule R applies to M iff SD(R) is not distinct from M.

Intuitively, then, the submatrix interpretation says 'the rule applies just in case there is an explicit match', while the distinctness interpretation says 'the rule applies just in case there is *no* explicit *mis-match*'.

A moment's thought will show that for completely specified matrices these two interpretations are equivalent. For example, if the only features available are [f] and [g], and every representation is specified for both these features, then if SD(R) is a submatrix of M, they are trivially nondistinct. In addition, if SD(R) = [-g] is nondistinct from some matrix M = [+f, -g], this can only be if there is a specific match between them. The difference between the two interpretations of application comes out only in cases where the representation is *unspecified* for a feature which is specified in SD(R). For example, given the SD(R) = [-g], and the representation M = [+f], then SD(R) is not a submatrix of M, but they *are* nondistinct. In this case, the rule would apply under the distinctness interpretation, but fail to apply under the submatrix interpretation.

It is of interest to note that SPE (p.336) adopts the distinctness interpretation of rule-application

Two units U<sub>1</sub> and U<sub>2</sub> are distinct if and only if there is at least one feature F such that U<sub>1</sub> is specified for [ $\alpha$ F] and U<sub>2</sub> is specified for [ $\beta$ F] where  $\alpha$  is plus and  $\beta$  is minus (p.336)

A rule of the form  $A \rightarrow B / X \text{ \_\_\_\_\_\_ } Y$  applies to any string  $Z = \dots X' A' Y' \dots$ , where  $X'$ ,  $A'$ ,  $Y'$  are not distinct from  $X$ ,  $A$ ,  $Y$  respectively... (p. 337)



Thus SPE adopts a distinctness interpretation of rule application, with respect to P-rules. As we have seen, however, the distinctness and the sub-matrix interpretations are equivalent in the case of completely specified matrices, and the grammar of SPE, explicitly adopting Stanley's arguments, is organized in such a way that P-rules only apply to such matrices.

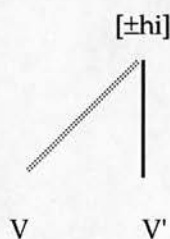
What is notable is that researchers in underspecification theory, in which P-rules do have access to incompletely specified matrices, have *not* adopted this interpretation of rule application, but (tacitly) the sub-matrix interpretation. This will be obvious from an inspection of Steriade's account of Montañes cited above, a representative example of underspecification analysis.

Recall that Steriade argues that the neutrality of /a/ — which fails to trigger [-high] harmony — is correlated with the fact that the [-high] specification is redundant just in the case of [+low] vowels:

The lack of trigger or blocker behaviour of low vowels in the height harmony of Pasiego indicates that these vowels lack height values at the stage when harmony applies. That low vowels cannot be underlyingly associated with any value for [high] follows from the fact that they represent the R-class of segments for [high]

The underlying representation of /a/ will simply be [+low], unspecified for height. The harmony rule may be formulated as:

*Height Harmony*



Crucially, this rule is interpreted as failing to apply to the matrix [+low], even though SD(R) and M are nondistinct — clearly a sub-matrix interpretation of rule application. Informally, if a rule can't explicitly 'see' a specification, it won't apply.

Now in a unification-based framework, the only interpretation of what it means for a representation to meet the structural description of a rule is in effect a non-distinctness interpretation. As we have seen in this theory there is no distinction between 'rule' and 'representation': the question is one of what it means for two feature structures to UNIFY; and one feature structure will unify with another 'unless they have conflicting information' (Shieber 1986:18): that is, there is *no* explicit *mismatch* of information. To use the same example, the unification of [-g] and [+f] succeeds, and is simply the feature structure [-g, +f].

Thus a unification-based approach to phonology cannot be sensitised to the redundant/non-redundant status of features through underspecification, since this presupposes a submatrix-interpretation of application.

#### 9.5.7. *Rule-application II: ordered application*

In underspecification theory, then, the identification of classes which are 'distinctively [+F]' assumes that rules apply under a sub-matrix interpretation, and that segments which are 'redundantly [+F]' lack the [+F] specification. How then can a rule refer to the natural class of segments which are simply '[+F]'? The theory makes crucial use of RULE-ORDERING.

In order for a phonological rule to 'see' a redundant value, that value must be *inserted* by a redundancy rule ordered to precede the application of the relevant phonological rule. As we have seen, however, in a constraint-based approach, rule-ordering is not an option. Thus what is required is some way of encoding redundancy representationally. This is what the theory of structured specification supplies. Moreover, this position — one that is forced on us by the theory-internal requirements of constraint-based approach — makes different predictions to the rule-based encoding, predictions which we will see in the next chapter are in fact correct. Thus the use of blanks presupposes a rule-based approach to redundancy, whereas the alternative supplies a representational approach is to be preferred. Once again we see that underspecification should be rejected, and replaced with the alternative, structured specification.

## 10. The non-derivational approach to specification

---

### 10.1. Introduction

As we discussed in the previous chapter, current accounts of the interaction between phonological rules and feature specification are crucially derivational. Predictable and default values are first eliminated from representations and then reintroduced by rule. In this, the UNDERSPECIFICATION account, the conditioning effect is achieved by virtue of the different degree of specification that is in place at the various stages of the derivation at which a particular rule applies. Early in the derivation, representations will be underspecified for redundant and default values, and the structural description of rules applying at this stage of the derivation — by virtue of the sub-matrix interpretation of rule application — will fail to match the underspecified representations. We have characterized this underspecification account informally in the following way: *features are invisible*. At the stage when a feature value is invisible in certain segments (the segments for which it is a redundant or default value) a rule which mentions that feature value will fail to apply to that segment: the invisibility of the feature specification allows the segment to escape the purview of the rule. The effect of a redundancy rule is precisely to make a value ‘visible’, so that, were the very same rule now to reapply to the representation, a different class of segments would match the structural description.

This metaphor allows us to succinctly express the fundamental difference between the underspecification approach and the proposed STRUCTURED specification account. Structured specification may be characterized in the following way: *rules are blind*.<sup>1</sup> Under this approach, the fact that a rule is sensitive to the redundant-distinctive status of a feature is a parameter of the rule itself. Under structured specification, representations encode the fact that certain values are redundant and that certain values are default values,

---

<sup>1</sup> A more exact analogy would be ‘some features are chameleons’ (underspecification) and ‘some rules are colour-blind’ (structured specification); ‘invisible’ and ‘blind’ will suffice to make the point, however.

through the *structural relations* that hold between the features of the representation. This encoding remains constant throughout the derivation. It is not modified in any way by 'redundancy rules', and thus a representation does not present a different appearance to one rule rather than another. It is thus 'up to the rule itself' whether it chooses to see the specification structure or not. Some rules are sensitized to these structural relations, and can thus distinguish redundant from non-redundant values; other rules are not so sensitized. This is formally expressed in the structural descriptions of rules, some of which mention features in a particular specification relation, others of which mention features *simpliciter*.

In the underspecification account, a rule which spreads only distinctive features, say, looks identical to a rule which spreads all features. The variable results of the two rules are a function of the variable specification of the objects the rule operates on at different stages of the derivation. Under our account, there are no stages in the derivation where redundant features are missing. The source of the redundancy effects, then, must be localized in the rule. The structured specification account thus adopts a non-derivational approach to phonological interactions. The conditioning effect of specification is not reliant in any way on ordering phonological rules with respect to redundancy rules in such a way as to encounter the requisite type of representation. Indeed, in the structured specification account, there *are* no redundancy rules. These two accounts make different predictions. Below, we cite evidence from Yoruba and from Maasai that supports the proposed non-derivational account. Before we consider this evidence, however, we need to make two points in order to forestall possible criticisms.

Firstly, although there is no possibility of extrinsically ordering redundancy rules with respect to phonological rules in our approach, and while we do not advocate the extrinsic ordering of phonological rules which crucially refer to distinctive values prior to rules which don't, there *is* an intrinsic order between these two classes of rules.

And secondly, although we claim that the redundant/non-redundant conditioning effects are properly considered a parameter of rules themselves, it is nevertheless true that the behaviour of rules *is* conditioned by the variable system of redundancy in the representations. This variation, however, is not a variation in degree of specification through the course of a



derivation, but variation in the pattern of redundancy that holds in various phonological contexts.

We now consider each of these points in more detail, before returning to the empirical evidence that supports our central claim.

## 10.2. The precedence of rules affecting distinctive features

Just as a distributive lattice is a lattice with extra structural constraints, so, in the structured specification approach, a 'distinctive' or non-redundant feature is a *feature* with extra structural constraints. Anything true of lattices in general is also true of distributive lattices; likewise, anything we predicate of a feature in general will also be true of a distinctive feature. We may thus have rules which refer to features *simpliciter*, which make no reference to the distinctive/redundant status of the features they manipulate; we may also, however, have rules which confine themselves to features which have the extra structural configuration of distinctive features. Moreover, there is a natural intrinsic ordering over such rules. A rule which mentions extra structure—which has a richer structural description—is *more specific* than a rule which omits that structure. Such rules will be ordered by the elsewhere condition. To put it another way, the set of sounds for which a feature is distinctive will always be a subset<sup>2</sup> of the set of sounds which have the feature *simpliciter*. Thus in the canonical five vowel system, the set of sounds which are [+back] are {a, o, u}; but the set of sounds which are unpredictably [+back] are {o, u}. Thus there is a natural sense in which rules which are concerned with distinctive values are more specific than rules which pay no heed to distinctive status.

As has been often observed (Calder 1988, Sproat 1992, Andrews 1990) the pervasive nature of the subsumption relation in unification-based approaches to grammar permits an elegant and computationally efficient expression of elsewhere condition effects: rules subsume one another, and the system can be set up in such a way that the most specific rule consistent with a given representation applies to that representation.

---

<sup>2</sup> Not necessarily a proper subset.



This ordering generalization is reflected in rule-based specification accounts. In such theories, redundant features are added during the course of the derivation. Thus in the early stages, the only features which are present are distinctive features; at later stages, after the application of redundancy rules, the features are a mixture of distinctive and redundant, with no discrimination among them possible in the formalism. Thus rules manipulating exclusively distinctive features tend to precede rules affecting features *simpliciter*. In our approach, on the other hand, this behaviour follows from the very *structure of distinctiveness*; a structure coded in the specification hierarchy.

### 10.3. Polysystemic effects

Under our approach, we consider sensitivity to distinctiveness or markedness a parameter of specific rules. Nevertheless, there are still a number of cases where the surface pattern of regularity is conditioned by the system of redundancy in the *representations*. This is due to the fact that the system of redundancy is not homogeneous for the whole of any language. In different contexts, a specific sound will occur in paradigmatic contrast with different sets of sounds. For a given position, then, the features which are predictable will be different from some other position.

This notion is well known from work by Firth and his followers, who viewed language as POLYSYSTEMIC. Its most recent manifestation is in the autosegmental licensing approach of Goldsmith (1989). The consequence of this for the current framework, is that the effect of a rule or constraint which requires, say, that the features that it manipulates be distinctive, will have different surface effects in different environments.<sup>3</sup>

An extension of this principle, which does not seem to have been previously investigated, is the notion that the system of *defaults* is not homogeneous for the whole of any language, that the effect of a rule or constraint which requires that the features that it manipulates be marked features, will also have different surface effects in different environments.

---

<sup>3</sup> See Ao (1991) for an example of such an effect in Kikongo.

Such effects are impossible to express in a framework such as radical underspecification, which treats the specification of the marked/unmarked value of each feature as if the phonology were ‘monosystemic’, applying redundancy rules in context-free fashion, across the board. In the following section, we advance an argument indicating that defaults can indeed be polysystemic. The discussion will also serve to introduce a number of aspects of the non-derivational approach which will be crucial in our subsequent discussion of Maasai. An important component of the approach is the use of DEFAULT LOGIC to model the interaction of phonological defaults and implicational constraints.

10.3.1.        *ATR phenomena*

The Proto-Kwa<sup>4</sup> vowel system had ten vowels: a typical five vowel system cross-classified by an ATR contrast (Stewart 1971:204):

(2)			
	i / *ɪ		u / *ʊ
	e / ɛ		o / ɔ
		*ɜ / a	

The combinations \*[+high, –ATR] and \*[+low, +ATR] are historically unstable. Presumably the reason for this is an articulatory tension involving tongue root and tongue body: it is difficult to hold the tongue body down while advancing the tongue root, hence the loss of [+ATR, +low] vowels. It is also difficult to raise the body of the tongue while keeping the root retracted, hence the loss of [+high, –ATR] vowels. It also seems that the low vowel is less stable than the high vowels, and languages tend to lose this vowel first. In the following discussion we will confine our attention to low vowels. Whatever the cause, the lack of [+low, +ATR] vowels can be formalized as a feature co-occurrence constraint:

(3)        \*[+ATR, +low]

We interpret this constraint logically, as:

---

<sup>4</sup> This section draws heavily on Stewart (1971). The main languages comprehended by the classification Kwa are: Akan, Ewe, Yoruba, and Igbo.

$$\neg ([+ATR] \wedge [+low])$$

which is thus equivalent to the following material implications, by *modus ponens*:

$$[+ATR] \supset [-low]$$

$$[+low] \supset [-ATR]$$

The loss of such vowels is widespread in Kwa languages, and is particularly interesting because of the interaction of the co-occurrence restriction with the systems of ATR HARMONY in the various languages. The data reveal that the compensation for this loss is not homogeneous, but has resulted in a variety of surface alternations and neutralizations both across and within different languages.

Following the lead of Pulleyblank (1986) and Archangeli and Pulleyblank (1989), we will attribute the behaviour of the modern reflexes of \*[ɜ] to the differing default values for the features [low] and [ATR] in these languages. We will go further, however. In the radical underspecification approach they propose, the behaviour of default values is homogeneous across the language. What we suggest is that the system of defaults may be different in different parts of the phonology — specifically, in stems as opposed to affixes — and that one and the same phonological rule may have different effects in different contexts due to the parameterizing effect of these defaults. Defaults thus exhibit a polysystemic, as opposed to a monosystemic behaviour.

Like Archangeli and Pulleyblank, we assume that rules are sensitive to whether the values they are manipulating are marked values or default values. As previously discussed, in the radical underspecification account that they propose, the conditioning of phonological rules with respect to marked/unmarked status is implemented by omitting default values from underlying representations, and by applying redundancy rules ordered late in the derivation to supply the missing values. For the purposes of exposition, and in order to bring out more clearly the differences between the two approaches, I will adopt this underspecification convention for the moment. Subsequently, we will find it necessary to revise this assumption, and introduce a constraint-based approach to defaults, but for the moment the argument will not be affected either way. The essential thing to grasp in what follows is that as we systematically alter the default specifications for ATR and vowel height, these specifications will interact with harmony rules and

the feature co-occurrence constraint formulated above to yield a variety of surface patterns.

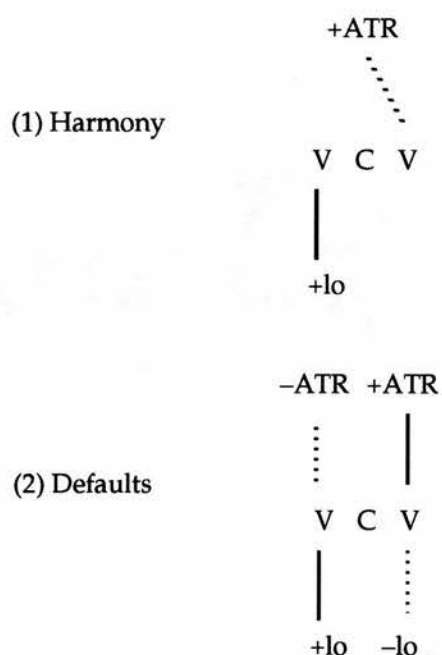
Consider a language which has a simple system of ATR harmony, in which all vowels within a stem share the same value for ATR. In such a language, entire morphemes may be characterized pre-theoretically as either [+ATR] or [-ATR]. We may represent the distribution of ATR as a morpheme-level floating feature, which associates and spreads to all vowels in its domain according to standard autosegmental principles. If, however, the language in question, like many in the Kwa group, lacks a [+ATR, +low] vowel, the spread of ATR will be complicated in morphemes which contain a low vowels.

10.3.2.        *Default setting 1: [ ] → [-ATR]; [ ] → [-low] (Asante)*

Consider first the case where the marked value for [ATR] is [+ATR], and the marked value for [low] is [+low]. In underlying representation, the low vowel of the language ([-ATR, +low]) will be simply [+low], with no ATR value specified. When [+ATR] attempts to dock on a vowel which is prespecified as [+low], the rule will be blocked by the feature co-occurrence restriction. A late default rule will subsequently insert [-ATR] across the board on segments unspecified for [ATR], and the [+low, -ATR] vowel will surface. Non-ATR morphemes, on the other hand, will have no floating specification in UR, and [-ATR] will be inserted across the board. This is exemplified in the following [VCV] stem, in which the first vowel a low-vowel, and the second is mid:

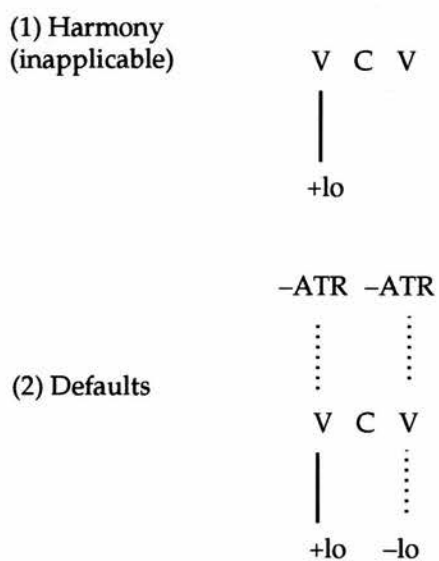
(4a)

*ATR morpheme*



(4b)

*Non ATR morpheme*



The result then is a suspension of the otherwise exceptionless rule of ATR-agreement, and we find 'mixed-roots' on the surface, stems with both [+ATR] and [-ATR] vowels, just in case the [-ATR] vowel is the low vowel.



Such a situation is exemplified by Asante, a dialect of Akan. In this language, in words containing no low vowels, all vowels belong either to the [+ATR] or [-ATR] set (Clements 1985:62):

<b>e-bu-o</b>	nest
<b>ɛ-bu-ɔ</b>	stone
<b>o-kusi-e</b>	rat
<b>ɔ-kɔdi-ɛ</b>	eagle

The low vowel, on the other hand, may occur in words containing vowels of either set:

<b>yari</b>	to be sick
<b>kari</b>	to weigh
<b>pɪra</b>	to sweep
<b>bisa</b>	to ask

Stewart (1971:199) observes:

The replacement of \*[ɜ] with [a] would appear to have occurred unconditionally in monosyllabic stems in Akan.

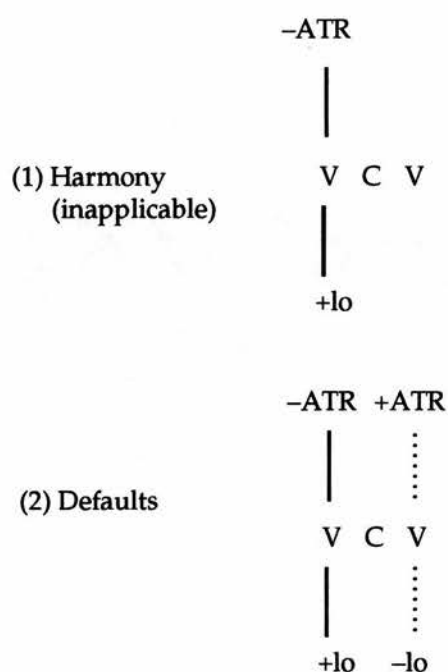
### 10.3.3. *Default Setting 2: [ ] → [+ATR]; [ ] → [-low] (Yoruba)*

Imagine now a second possibility: a language, otherwise similar to Asante, but for which, unusually, the marked value of [ATR] is [-ATR]. Archangeli and Pulleyblank (1989) argue that Yoruba is such a language. Here, the representation of /a/ underlyingly will be [+low] *and* [-ATR], and it will be the [-ATR] specification that spreads, with [+ATR] specifications being supplied by default.

When the [-ATR] specification spreads to a position occupied by a low vowel, explicitly marked as [+low], a [-ATR, +low] segment will result as a direct effect of the harmony rule, rather than through default specification. In [+ATR] words, there will be no floating feature, and [+ATR] will be filled in by default; but not, of course, on low vowels, which are already specified [-ATR], and for which the feature co-occurrence constraint in any case prevents the assignment of [+ATR] to [+low] vowels.

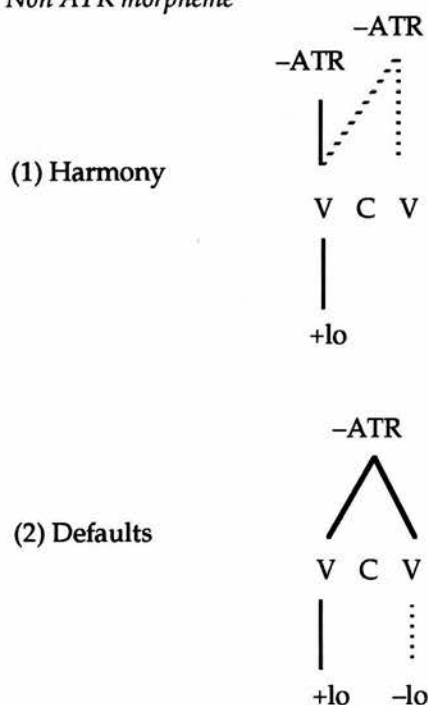
(5a)

*ATR morpheme*



(5b)

*Non ATR morpheme*



(We take no position on the vacuous spread of [-ATR] in these cases: see Archangeli and Pulleyblank (1989) for discussion).

Interestingly, then, we find the same surface behaviour with respect to low vowels whether [+ATR] or [-ATR] is taken to be the unmarked value, provided only that the marked value for low is [+low]. In the first case, Asante, the co-occurrence constraint blocks the assimilation rule; in the second case, Yoruba, the same co-occurrence constraint blocks default value assignments. In either case we have a vowel which fails to undergo [ATR] harmony, leading to mixed-vowel roots, exhibiting exceptions to the otherwise homogeneous distribution of ATR. In each case, where the impermissible [+ATR, +low] vowel would result through the unconstrained combinatorics of harmony and segmental specification, it is the [ATR] value that 'gives way', and the [low] value which remains invariant: [+low].

#### 10.3.4. *Default setting 3: [ ] → [-ATR]; [ ] → [+low] (Okpe)*

We now consider systems in which the marked value for [low] is [-low]. In such systems, the value [+low] is the unmarked value, the default value — intuitively, the value which surfaces unless there is 'good reason' to the contrary. Thus, in an underspecification approach, the low vowel will not in fact have a specification for [low] in UR. Given the default status of this feature, we expect to see some malleability with respect to its realization. We first consider such a system as it interacts with a harmony rule for which [+ATR] is the marked value for ATR. As before, the same feature co-occurrence constraint concerning ATR and vowel height is in place.

Before we proceed, it is important to consider the interaction of redundancy rules and default specifications. Note that the feature co-occurrence constraint rules out a combination of two properties; that is, it has the logical form  $\sim(p \& q)$ . On a constraint-based interpretation of redundancy, this is logically equivalent to the conjunction of implications:

$$((p \supset \sim q) \& (q \supset \sim p))$$

Thus, given the constraint \* [+ATR, +low], we also know that

$$\begin{aligned} [+ATR] &\supset [-low], \text{ and} \\ [+low] &\supset [-ATR] \end{aligned}$$

Consider now how these implications concerning vowel height and ATR interact with the defaults we postulate are now in place: the antecedent of the

first implication — [+ATR] — is a marked property, but the antecedent of the second — [+low] — is a default property.

In recent years, the properties of logical systems with defaults have been subject to a great deal of attention. As Nutter (1983:297) observes:

Reasoning from incomplete information and from default generalizations follows patterns which standard first order predicate logic does not reflect. The most striking deviation involves making inferences whose conclusions may be counterindicated by further information which does not explicitly contradict anything previously known

Reiter (1980) contains the following lucid account of some basic properties of such a nonmonotonic logic system (the reason for quoting this passage at such length will become apparent below):

A good deal of what we know about the world is 'almost always' true, with a few exceptions. Such facts usually assume the form "Most *P*'s are *Q*'s", or "Most *P*'s have property *Q*". For example most birds fly except for penguins, ostriches, the Maltese falcon etc. Given a particular bird, we will conclude that it flies unless we happen to know that it satisfies one of these exceptions. How is the fact that most birds fly to be represented? The natural first order representation explicitly lists the exceptions to flying:

(x).BIRD(x) & ~ PENGUIN(x) & ~ OSTRICH(x) & ...  $\supset$  FLY(x).

But with this representation we cannot conclude of a 'general' bird that it can fly. To see why, consider an attempt to prove FLY(tweety) where all we know of tweety is that it is a bird. Then we must establish the subgoal

~ PENGUIN(tweety) & ~ OSTRICH(tweety) & ...

which is impossible given that there is no further information about tweety. We are blocked from concluding that tweety can fly even though intuitively we want to deduce just that.

What is required is somehow to allow tweety to fly *by default*. How is this default to be interpreted? We take it to mean something like "If *x* is a bird, then in the absence of any information to the contrary, infer that *x* can fly". The problem is then to interpret the phrase "in the absence of any information to the contrary". The interpretation we adopt is "It is consistent to assume that *x* can fly". Thus "If *x* is a bird and it is consistent to assume that *x* can fly, then infer that *x* can fly". We represent this more formally by the following *default rule*:

$$\frac{\text{BIRD}(x):M.\text{FLY}(x)}{\text{FLY}(x)}$$

Here *M* is to be read as "it is consistent to assume". The exceptions to flight are given a standard first order

representation.

$(x).PENGUIN(x) \supset \sim FLY(x)$

$(x).OSTRICH(x) \supset \sim FLY(x)$

etc.

Notice that if  $FLY(tweety)$  is inferred by default then the assertion  $FLY(tweety)$  has the status of a *belief*; it is subject to change, say by the subsequent discovery that *tweety* is a penguin. We can then reinterpret the default rule as saying "If  $x$  is a bird and it is consistent to believe that  $x$  can fly then one may believe that  $x$  can fly".

Now this discussion of the logic of defaults has a direct application to phonology, provided we have given our phonological formalism a logical foundation. And as we have seen, this is precisely what a constraint-based approach to redundancy provides. To illustrate the applicability of these notions, consider Reiter's exposition again, but this time applied to a plausible set of phenomena from the phonological domain. We make the following substitutions:

For BIRD	read VOWEL
For FLY	read LOW
For PENGUIN	read HIGH
For OSTRICH	read ATR
For <i>Tweety</i>	read /a/

A good deal of what we know about a sound system is 'almost always' true, with a few exceptions. Such facts usually assume the form "Most *P*'s are *Q*'s", or "Most *P*'s have property *Q*". For example most vowels are low except for high vowels, and say, in some particular system, [+ATR] vowels. Given a particular vowel, we will conclude that it is low unless we happen to know that it satisfies one of these exceptions. How is the fact that most vowels are low to be represented? The natural first order representation explicitly lists the exceptions to height:

$(x).VOWEL(x) \& \sim high(x) \& \sim +ATR(x) \& \dots \supset low(x)$ .

But with this representation we cannot conclude of a 'general' vowel that it is low. To see why, consider an attempt to prove  $low(a)$  where all we know of /a/ is that it is a vowel. Then we must establish the subgoal

$\sim high(a) \& \sim +ATR(a) \& \dots$

which is impossible given that there is no further information about /a/. We are blocked from concluding that /a/ is low even though intuitively we want to deduce just that.

What is required is somehow to allow /a/ to be low *by default*. How is this default to be interpreted? We take it to mean something like "If  $x$  is a vowel, then in the absence of any information to the contrary, infer that  $x$  is low". The problem is



then to interpret the phrase "in the absence of any information to the contrary". The interpretation we adopt is "It is consistent to assume that  $x$  is low". Thus "If  $x$  is a vowel and it is consistent to assume that  $x$  is high, then infer that  $x$  is low". We represent this more formally by the following *default rule*:

$$\frac{\text{VOWEL}(x):M. \text{low}(x)}{\text{VOWEL}(x)}$$

Here  $M$  is to be read as "it is consistent to assume". The exceptions to height are given a standard first order representation.

$$(x).\text{high}(x) \supset \sim \text{low}(x)$$

$$(x).+\text{ATR}(x) \supset \sim \text{low}(x)$$

etc.

Notice that if  $\text{LOW}(a)$  is inferred by default then the assertion  $\text{LOW}(a)$  has the status of a *belief*; it is subject to change, say by the subsequent discovery that  $/a/$  is  $[+\text{ATR}]$ . We can then reinterpret the default rule as saying "If  $x$  is a vowel and it is consistent to believe that  $x$  is low then one may believe that  $x$  is low".

It is apparent that these notions can be translated directly into the phonological domain, and armed with these notions we can provide a logical interpretation of default behaviour without recourse to underspecification and the derivational interpretation. It will be observed that, in terms of our previous discussion of these matters, Reiter's logic expounds a 'credulous' interpretation of default behaviour, in which segments with default features are interpreted as if they really did possess those features, unless and until this belief is counterindicated. We could imagine, however, an alternative 'skeptical' logical system, in which segments with default features were interpreted as lacking the relevant features. That is, no deduction is built upon such features unless and until they are corroborated.

In the present instance, it is enough to note that, logically, defaults are inherited. If we know that  $p$  implies  $q$ , and that some segment  $S$  is  $p$  by default, then we can assume it is  $q$  by default. Nutter (1983:298) describes such default inference in the following way:

The guarded status of default generalizations must be inherited through inferences, and from a guarded proposition, it must be possible to infer a guarded version of any conclusion which could be inferred from the unguarded version of the premise.

In other words, it is consistent to assume  $S$  is  $q$ , just as long as this default expectation is not counterindicated by some fact revealing that  $S$ , contra that

expectation, is not in fact *p*. Thus implications built on the back of defaults inherit the caveat that they are no stronger than the foundation they stand upon.

Now the first half of the conjunction above — the logical implication  $[+ATR] \supset [-low]$  — is not hedged by defaults. This follows because the specification  $[+ATR]$  is not a default specification. Thus this constraint holds without proviso, and we know of any segment which is  $[+ATR]$ , that it is also, redundantly,  $[-low]$ . The second half of the conjunction, however —  $[+low] \supset [-ATR]$  — does interact with the defaults. Recall that we are assuming that objects which are  $[+low]$  are so only by default. The implication  $[+low] \supset [-ATR]$  is founded on a default assumption, and thus inherits the guard associated with that default. Thus the status of  $[-ATR]$  for such segments is also, effectively, a default specification: it may be counterindicated by positive evidence that the segment in question is not, in fact, what the defaults lead us to assume. The significance of this is that in the following derivation we will want the implication  $[+ATR] \supset [-low]$  — the unguarded implication — to take precedence over the default expectation that, other things being equal, segments are  $[+low]$  — the default setting for that feature. The discussion above shows that the precedence we seek is one of logical precedence, and obviates the need to order a blank-filling redundancy rule before a blank-filling default rule.

To return to our main argument, then, consider then the spread of  $[+ATR]$  harmony to a low vowel in a system where  $[+low]$  and  $[-ATR]$  are the default specifications. In radical underspecification terms, the low vowel will be unspecified for *both*  $[ATR]$  and  $[low]$ , since its values for these features will be the unmarked values  $[-ATR]$  and  $[+low]$ . In a  $[+ATR]$  morpheme, then, the  $[+ATR]$  specification will spread to the empty position:

(6a.i)

*ATR morpheme*

(1) Harmony



The spread of [+ATR] is not *blocked* by the co-occurrence constraint, but it *is* conditioned by it: as we have seen, the constraint that [+ATR] and [+low] are incompatible implies that [+ATR] objects are [-low]. Thus as an interaction of the assimilation rule and the co-occurrence constraint, we have:

(6a.ii)

*ATR morpheme*

(2) FCR

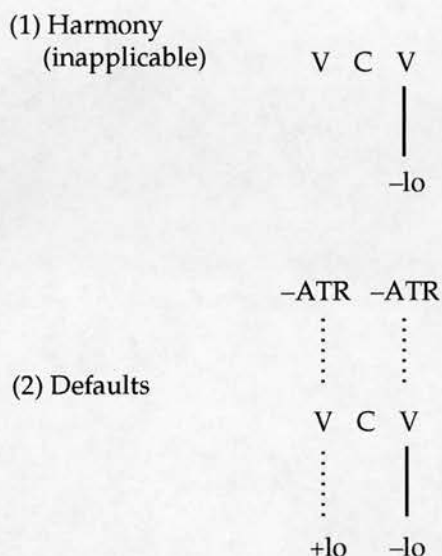


As we have seen, the implication that the initial vowel is [-low] takes logical precedence over the default statement that vowels are [+low]. The result, then, is that the default low vowel will *raise* in this context. Rather than the incompatibility of the feature values [+ATR] and [+low] being resolved in favour of [+low], leading to an interruption of the harmony stretch, the incompatibility is resolved in favour of [+ATR], preserving the integrity of the harmony domain, but forcing a concomitant modification in vowel height.

In a non-ATR morpheme, on the other hand, [-ATR] will be inserted by default. This, of course, is perfectly compatible with the default specification for the low vowel, and thus, there being no good reason to the contrary, the low vowel will surface by default:

(6b)

*Non ATR morpheme*



What we expect to see in such a system, then, is a low-vowel/mid-vowel *alternation*, and no interruption to the harmony domain. Just such a case is found in Okpe (Pulleyblank 1986):

<b>so</b>	sing
<b>ti</b>	pull
<b>ε-so</b>	to sing
<b>e-ti</b>	to pull
<b>a-so-ri</b>	we sang
<b>e-ti-ri</b>	we pulled

Here, the infinitive marker /E/ exhibits a simple [+ATR]/[-ATR], [e / ε] alternation controlled by the [ATR] specification of the stem. However, the prefix associated with the past morphology exhibits an [e / a] alternation: that is, in [-ATR] environments the low vowel surfaces unmodified, whereas in [+ATR] environments it both undergoes harmony and raises.

The strategy employed here is extremely important for the constraint-based approach to phonology. A common objection to thorough-going constraint-based approaches is that the mere statement that some constraint holds is insufficient to determine the particular strategy a language might choose to meet that constraint. It is this that has led many theorists to prefer thoroughly

rule-based accounts, or to adopt constraints and introduce concomitant 'repair strategies' or implementation rules, which are 'triggered' by constraint violations. What we see here is a single constraint \*[+low, +ATR] at work in a variety of languages, but with different resolutions. However, it is unnecessary to write an explicit rule to handle the variation. Rather, the particular resolution in each case is the only one which can simultaneously satisfy the combination of logical constraints and defaults focussed on a particular position. Outputs are here parameterized by defaults.

What we have seen so far, then, is a variation in the pattern of compensation for the loss of the proto-vowel \*[ɜ] in different languages, which is accounted for solely by the interaction of a common co-occurrence constraint on the one hand, and the different settings for default specifications on the other.

Suppose now, we find exactly the same variation of compensation occurring in different environments in the *same* language. For example, disharmony in roots accompanied by alternation in affixes. Or, disharmonic prefixes, with alternating suffixes. This would provide *prima facie* evidence for different default settings in different parts of the morphophonology. Such cases are in fact attested. Recall Stewart's generalization quoted above:

The replacement of \*[ɜ] with [a] has occurred unconditionally in monosyllabic stems in Akan

We have seen an example of this pattern in the Asante dialect. However, Stewart (1971:200) goes on to observe:

In the Fante dialect of Akan, prefixes which have a [a] before [–ATR] stems have not \*[ɜ] but [e] before stems with [+ATR] vowels

The pattern in Fante is thus one of neutralization in stems, and alternation in prefixes: in a sense, it is an Asante language in its stems, and an Okpe language in its prefixes. We account for the variation *within* the language just as we do between languages. The pattern of defaults is not homogeneous throughout the language. What we have is a polysystemic system, but one based not on different patterns of redundant information, but different patterns of default information.

#### 10.3.5.      *Default settings 4: [ ] → [+low]; [ ] → [+ATR] (Akuapem)*

There is a final combination of default specifications which we have not considered: and this combination has certain theoretical peculiarities. This is



the pattern of defaults in which the default specification for low is [+low] and the default specification for ATR is [+ATR]. The peculiarity here is that this is precisely the combination which the co-occurrence constraint we have been considering rules out.

There are two responses we might make with respect to such a situation. On the one hand, we might argue that such a system is logically inconsistent, and set things up in such a way as to make such a default specification impossible. In other words, we could set up the default logic in such a way that the statement 'x is p by default and 'no x is p' is a formal contradiction.

On the other hand, we may hypothesize that phonological systems are of such a nature that they *do* permit such clashes of constraints and defaults, and seek empirical examples corresponding to such a theoretical possibility. We might reason as follows: We suppose it must be a universal fact that if a default value cannot be instantiated due to a co-occurrence constraint, the opposite value must be realized. Now the default value [+ATR] cannot be instantiated due to the constraint preventing its co-occurrence with [+low], and similarly the default value [+low] cannot be instantiated due to the constraint preventing its co-occurrence with [+ATR]. This is admittedly an extraordinary situation. Two unordered defaults are striving to instantiate the same object, but each blocks the other. The logical outcome of such a situation might plausibly be that *neither* default is activated; that is, that a vowel should surface that is *neither* [+low] *nor* [+ATR] — even though it occurs in an [+ATR] environment, and even though it alternates with a low vowel.

Interestingly, just such vowels occur in ATR harmony systems, and there is no coherent account of them. Here is Stewart (1971:201) again:

In the Akuapem dialect of Akan [ɜ] has evidently been replaced by [ɛ]; Schachter (1962:7) reports that prefixes which have [a] before stems with root-unadvanced vowels have not [ɜ] but [ɛ] before stems with root-advanced vowels irrespectively of whether the first vowel of the stem is front or back

Thus we find a vowel, which alternates with [a] (a [+low] vowel) and occurring in a [+ATR] environment, which nevertheless surfaces as both [-low] and [-ATR].

In fact such vowels also occur in *suffixes* in Okpe. Omamor (1988:50) observes:

The fact that [a] alternates with [ɛ] as its [+ATR] counterpart introduces some complication into the otherwise neat system of vowel harmony in Okpe

In the current state of our knowledge, we must admit that intuitions give out here, both with regard to the behaviour of phonological systems, and with respect to the logical interpretation of default reasoning. Nor do we wish to underestimate the difficulties of teasing apart phonetically the distinctions which form the basis of the phonological argument (Hess 1988; Lindau-Webb 1988). Accordingly, we leave open the possibility of a correlation between the fourth combinatorial category and this intriguing vowel: a vowel which is the [+ATR] harmonic counterpart of a [+low] vowel, but which is neither [+ATR] nor [+low].

#### **10.4. Against derivational specification**

We now return to our central argument, and present two cases which indicate the derivational approach to specification advocated by underspecification theory is incorrect, and that the non-derivational approach provides a solution to the problematic cases.

##### *10.4.1. Yoruba*

Clements and Sonaiya (1989) present a number of arguments against the radical underspecification approach to Yoruba phonology adopted by Pulleyblank (1988). In Pulleyblank's account, a number of asymmetries in the behaviour of the vowel [i] are explained in terms of its unique status as the DEFAULT VOWEL in the language, unspecified for any place features in underlying representation. One such asymmetry is the following. Across words, Yoruba has an optional rule of regressive assimilation, whereby the last vowel of a word completely assimilates to the following vowel in the next stem, unless the stem initial vowel is /i/. Pulleyblank expresses this in terms of a simple rule of place feature assimilation. The rule fails to spread any place feature at all in the case of /i/, since all its place features are default features, and thus unspecified.

By way of criticism, Clements and Sonaiya (1989) motivate various rules and constraints which need crucially to identify /i/ at a stage of derivation where its lack of specifications would make this impossible.

For example, Yoruba contains the following phonotactic constraint. In [VCV] nominal stems, if the first vowel is [a] or [ɔ], the second cannot be [i] or [u]. That is:

$$* [\{a, \text{ɔ}\} \text{ C } \{i, u\}]_N$$

Clements and Sonaiya (1989:11) formulate the constraint as follows:

$$* [ [-\text{ATR}, +\text{back}] \text{ C } [+ \text{high}, -\text{nasal}] ]_N$$

They conclude:

Clearly, in order to state this constraint, which holds over stems at level 1, we must assume that the vowel [i] bears the feature [+high]; if it did not, the constraint would no exclude roots and stems in which the second vowel was [i].

The problem, then, is that certain rules and constraints, such as those discussed by Pulleyblank (1988), treat [i] as if it lacked its default specifications, even as late as the post-lexical phonology — the point at which vocalic assimilation rule mentioned above takes place — whereas other rules and constraints, such as this one discussed by Clements and Sonaiya, treat [i] as if its default values were present, even as early as level 1 morphophonology. Since values cannot be simultaneously present and absent, it is impossible to account for the behaviour of [i] in Yoruba.

This problem results from attempting to account for the conditioning effects of specification by modifying the degree of specification derivationally. Intuitively, what the Yoruba case exemplifies is simply one in which certain rules and constraints are sensitive to the marked-unmarked distinction, while certain constraints are not. The rules discussed by Pulleyblank manipulate only marked values, the constraint discussed by Clements and Sonaiya refers to values irrespective of their status. A rule which is parameterized to spread, not just vowel features, but vowel features in the particular structural relation that marked specifications bear, will thus not affect the specifications of [i], all of whose place features are unmarked; but a constraint which is parameterized to rule out combinations of features *simpliciter* will effect [i] along with all other vowels. A structured specification approach thus predicts precisely the kind of behaviour exhibited in Yoruba, where a segment behaves one way with respect to some rules, and another way with respect to others. The explanation is simply that some *rules* are of a type that hold of only marked features; other rules are of a type that hold of all features,

marked and default. It is the rule that is the locus of the variability, not the representation.

Mohanan (1991) notes that the facts of English assimilation are similarly problematic. We refer the reader to this paper for details of the analysis, and simply quote Mohanan's (1991:315) conclusion:

radical underspecification claims that from the point in the derivation where the unspecified value is specified, the distinction between non-dominant and dominant will cease to be operative [...] the facts of English assimilation constitute a counterexample to this prediction. The derivation in the lexical module crucially requires reference to the non-dominant feature value ([+cor]), and yet this value continues as non-dominant in the derivation in the post-lexical module<sup>5</sup>

The Yoruba example, although adequate to introduce the general point, is not totally satisfactory. As we have seen, in our approach, phonotactic constraints are stated quite differently from the form assumed by Clements and Sonaiya. For us, the fact that there is an implicational dependency between [ATR, back] at V<sub>1</sub> and [high, nasal] at V<sub>2</sub> will be expressed through the relative position of these features in the lattice of representations. The constraint is immanent in the pattern of storage, rather than stated in isolation from the lexical forms which expound it. Comparisons with the standard model are thus difficult to draw.

However, we will now turn to a second case, in which the comparisons are easy to make, and which is in any case more persuasive. The kind of phenomenon which would clinch the argument that the sensitivity we have been discussing is a parameter of specific rules rather than the degree of specification at a certain stage in the derivation would be a case in which one and the same segment in a particular representation, simultaneously within the scope of two distinct rules, behaved as if its default value for some feature [F] were visible to one rule, but not to the other. In such a case, we would be compelled to say that indeed, one rule was blind.

---

<sup>5</sup> Mohanan also notes the problematic nature of the Yoruba data.



Just such a case is found in Maasai<sup>6</sup>, where one and the same segment is OPAQUE with respect to one rule of ATR harmony, but TRANSPARENT with respect to another. The opacity/transparency effect is convincingly accounted for by the sensitivity of the rules to marked versus unmarked information. But since one and the same morpheme cannot simultaneously have two representations — cannot be simultaneously underspecified and fully specified — it follows that the different behaviour of the morpheme in the face of the two rules must be a parameter of the rules themselves.<sup>7</sup>

#### 10.4.2. *Maasai*

In the examples of alternations between low and mid vowels in ATR-harmony systems presented so far, the low vowel has alternated with a *front* mid vowel. Given that the low vowel is not distinctively front, however, it might be expected that we find alternations with *back* mid vowels in certain languages. Indeed, we would expect this possibility to be attested in cases where the default specification for [back] is [+back]. Such a case is found in Massai, to which we now turn. More importantly, the Maasai data provide conclusive evidence that sensitivity to marked/unmarked information must be coded in rules, rather than representations.

In Maasai /a/ does not undergo leftward [+ATR] harmony, surfacing as [a], but *does* undergo rightward [ATR] harmony, surfacing as [o] (the back version of the mid [+ATR] vowel, rather than the front version [e], as in Okpe prefixes). Following the line of argumentation we have been developing, we will conclude that these different behaviours are to be accounted for in terms of default specification. But rather than a situation in which a single rule

---

<sup>6</sup> The problematic nature of the Maasai data was noted in Churma (1988). See also Archangeli and Pulleyblank (in press).

<sup>7</sup> It might be argued that a redundancy rule could be extrinsically ordered so as to 'split' the application of the two ATR rules. Underspecification theorists, however, have properly eschewed such ad hoc extrinsic orderings between phonological rules and redundancy rules. Rather, they have attempted to derive the relative ordering of redundancy rules from general principles. The standard position is that all the redundancy rules which are applicable at any level of the lexical phonology are applicable throughout that level (see, for example, Pulleyblank 1986:125). Thus the only principled way in which the two ATR rules could encounter differently specified representations would be if it could be independently argued that the two rules applied at different levels of the lexical phonology. As we will see, this cannot be maintained in the Massai case.



produces different outputs by virtue of encountering different representations, we have here a case where two different rules, both of which simply spread [+ATR], produce different outputs, even when encountering the same representation. It might be argued that the /a/ which the leftward rule encounters, and the /a/ which the rightward rule encounters, are in fact differently specified underlyingly. Such could be maintained, for example, if the leftward rule only encountered prefixes, and the rightward rule only encountered suffixes. We might then argue that the difference arises from the different pattern of defaults operating in prefixes and suffixes. However, it can be shown in Maasai that this is not the case.

We claim that the two different behaviours of /a/ with respect to the ATR-harmony rules is precisely the behaviour we would expect if one rule is sensitive to the fact that the features it is manipulating are default features, and the other rule is insensitive to this dimension. Rather than one and the same rule having different effects in different morphophonological domains, this is an example of two rules, which spread the same feature, having different effects in one and the same domain. The different effects of the rules can be explained by the single parameter that one is sensitive to markedness distinctions, and one is not.

Maasai exhibits a pattern of dominant/recessive harmony in which the active value is [+ATR]. In dominant/recessive systems, morphemes belong to one of two classes. One class of morphemes, the dominant set, always bear a fixed value for [ATR]; the second class, the recessive set, alternate with respect to their value for [ATR]. Moreover, there is a correlation with respect to the distribution of dominance and [ATR]: all dominant morphemes have the same value for [ATR] — in Maasai, all dominant morphemes are [+ATR]. If any word contains a dominant morpheme, then the entire word is [+ATR]. A word is [–ATR] if and only if it consists solely of recessive morphemes. Such a language provides the classic motivation for default underspecification. Dominant morphemes are lexically specified as [+ATR], and this value spreads wherever it occurs: [–ATR] is not specified lexically, but is filled in by default late in the derivation. Accordingly, a recessive morpheme will surface as [–ATR] just in case there is no dominant morpheme in the word. Finally, the dominant/recessive distinction bears no relation to morphological categorization: both affixes and roots may be dominant, both affixes and roots may be recessive.

Like Yoruba /a/, Maasai /a/ exhibits blocking effects with respect to [ATR]:

- |      |                       |            |
|------|-----------------------|------------|
| (7a) | ɪ - [tɔn]             | 2-sit      |
| (b)  | i - [ton] - <b>ie</b> | 2-sit-APPL |
| (c)  | ɪ - [as] - <b>ie</b>  | 2-do-APPL  |

The applicative morpheme [**ie**] is dominant (we will show dominant morphemes in bold type, and roots will be bracketted). Note that by default — i.e. when there is no dominant morpheme in the word — all morphemes surface as [−ATR] (7a). When a dominant morpheme is concatenated with recessive morphemes (7b), the marked [+ATR] specification spreads — here leftward — to all morphemes in the word. However, the low vowel of the root does not undergo harmony (7c), and blocks the spread of harmony leftward. Like Yoruba, Maasai has lost its [+ATR, +low] vowels, and the failure of /a/ to undergo harmony can be expressed through the same co-occurrence constraint.

This then is an example of /a/ in a root blocking harmony. If /a/ occurs in a prefix, it exhibits the same behaviour — it fails to undergo harmony and blocks the spread of harmony:

- |      |         |         |                  |            |                    |
|------|---------|---------|------------------|------------|--------------------|
| (8a) | ki      | - [dot] | - un - <b>ie</b> | 3PL        | - pull - MT- APPL  |
| (b)  | kɪ - ta | - [dot] | - un - <b>ie</b> | 3PL - PAST | - pull - MT - APPL |

However, if /a/ occurs in a suffix following a [+ATR] stem, it both undergoes harmony and raises, surfacing as [o], and allows the harmony to spread:

- |      |                      |              |               |                   |
|------|----------------------|--------------|---------------|-------------------|
| (9a) | aa - [nk]            | - ɪ - ta - ɪ | 1S - nauseate | - ? - CONT - PASS |
| (b)  | aa - [ <b>ipot</b> ] | - i - to - i | 1S - call     | - ? - CONT - PASS |

Here the dominant root [**ipot**] spreads its [+ATR] specification rightwards: the continuative morpheme, which surfaces as [ta] in the [−ATR] word, undergoes [+ATR] harmony, surfacing as [to] and does not block the spread of [+ATR] rightwards.

At this stage, we might propose an account similar to that for Okpe. We might propose that the default specifications for *roots* and *prefixes* are such as to resolve the [+ATR, +low] incompatibility by preserving the marked [+low] specification, and blocking [+ATR]; while the defaults in *suffixes* are such as to resolve the incompatibility by preserving [+ATR] and overriding the default [+low] specification. However, in Maasai low-vowelled suffixes may be surface *either* as [a] or [o]:

(10a)	ε - tu	- [nuk]	- ar - ie	3S - PAST - bury - MA - APPL
(b)	e	- [ibuk]	- or - ie	3S - pour - MA - APPL

It is clear that the suffix /Ar/ does not undergo harmony *from the right* (10a), but does undergo harmony *from the left* (10b). The two rules 'see' one and the same morpheme differently. The crucial thing is whether or not the rule sees the [+low] specification on the /a/: if it does, the attachment of [+ATR] is blocked by the co-occurrence constraint; if it doesn't see the [+low] specification, the [+ATR] attachment takes place, with concomitant raising, preempting the default. Now the two rules look at the same object, but see different things. Thus one and the same morpheme is opaque with respect to one rule spreading [+ATR], and transparent with respect to another spreading [+ATR].

We account for this behaviour in the following way. The first rule says simply 'Spread [+ATR] leftwards'. This rule sees a segment with the specification [+low], and is 'unaware' whether this specification is a default specification or not. In fact, this rule doesn't even know there is a distinction between default specifications and marked specifications. To such a rule, this is just a [+low] vowel *simpliciter*. And to such a rule, the spread of [+ATR] is interdicted by the co-occurrence constraint. This is analogous to a rule which is unaware that there is a distinction between distinctive and redundant values. It deals simply with properties, and is not conditioned by the default/non-default or the distinctive/non-distinctive status of these features within the phonological system of the language.

The second rule has a crucial difference. In a similar way to the first, it says 'Spread [+ATR] rightwards'. But this rule is alive to the default/non-default status of the features it manipulates: it is conditioned by the systemic status of these features within the phonology. This rule sees a segment with the specification [+low], but it also sees that this is a default specification. As such, the default value may be preempted by the attachment of [+ATR] without contradiction.

A special quality of this case is that the default specification is permissive, rather than non-permissive. Usually, predictable and default specifications are 'inert': they fail to undergo, or fail to trigger a particular rule. The lack of specification usually entails that representations fail to meet the structural description of the rule. In this case however, the driving force behind the

analysis is not a rule but a *filter*, a feature co-occurrence constraint. Here, the default status of the specification fails to trigger the filter, and is thus *enabled* to undergo a phonological process. Default specification, or default sensitivity, is not in this case simply conditioning how a rule applies to a form. There is an extra dimension: it is conditioning how a constraint should be invoked to monitor the application of a rule to the target representation.

The precise formalization of the 'structural description' of this rule depends in part on just how the default/non-default distinction is encoded in the lattice of representations. This, as we have seen, remains an unresolved question. The purpose of the above discussion has been to investigate some of the properties of default behaviour that any formalization will have to deal with, and to establish that such a formalization must needs focus on rule-parameterization, rather than the derivational modification of representations. In addition, a non-derivational approach to specification must encode default structure in representations themselves, and thus provides support for the programme of structured specification.



## Bibliography

---

- Allen, W. (1951). Some prosodic aspects of retroflexion and aspiration in Sanskrit. *Bulletin of the School of Oriental and African Studies* 13:939–946.
- Allen, W. (1957). Aspiration in the Harauti nominal. In *Studies in Linguistic Analysis* [Special volume of the Philological Society]. Blackwell: Oxford.
- Allen, W. (1973). *Accent and Rhythm: Prosodic features of Latin and Greek*. Cambridge: C.U.P.
- Anderson, S. (1969). *West Scandinavian Vowel Systems and the Ordering of Phonological Rules*. PhD dissertation: MIT.
- Anderson, S. (1970). On Grassmann's Law in Sanskrit. *Linguistic Inquiry* 1:387–396.
- Anderson, S. (1985). *Phonology in the Twentieth Century: Theories of rules and theories of representations*. Chicago: University of Chicago Press.
- Andrews, A. (1990). Unification and morphological blocking. *Natural Language & Linguistic Theory* 8:507–558.
- Ao, B. (1991). Kikongo nasal harmony and context-sensitive underspecification. *Linguistic Inquiry* 22:193–196.
- Archangeli, D. (1984). *Underspecification in Yawelmani phonology and morphology*. PhD Dissertation, MIT. [Published New York: Garland, 1988].
- Archangeli, D. (1988). Aspects of underspecification. *Phonology* 5:183–207.
- Archangeli, D., and D. Pulleyblank. (1989). Yoruba vowel harmony. *Linguistic Inquiry* 20:173–217.
- Archangeli, D., and D. Pulleyblank. (in press). *Grounded Phonology*. Cambridge, Mass.: MIT Press.
- Bird, S. (1990). *Constraint-Based Phonology*. PhD dissertation: University of Edinburgh.
- Bird, S. ed. (1991). *Declarative Perspectives on Phonology* [Edinburgh Working Papers in Cognitive Science, Vol. 7]. Centre for Cognitive Science, University of Edinburgh.
- Birkhoff, G. (1948). *Lattice Theory*. Providence, R. I.: American Mathematical Society.
- Bresnan, J. ed. (1982). *The Mental representation of grammatical relations*. Cambridge, Mass.: MIT Press.
- Bromberger, S., and M. Halle. (1989). Why phonology is different. *Linguistic Inquiry* 20:51–70.
- Brown, G. (1969). Syllables and redundancy rules in generative phonology. *Journal of Linguistics* 6:1–17.
- Butler, C. (1985). *Systemic Linguistics: theory and applications*. London: Batsford.
- Calder, J. (1988). Paradigmatic morphology. In *Association for Computational Linguistics Proceedings, 4th European Meeting*.



- Cherry, E. C., M. Halle, and R. Jakobson. (1953). Toward the logical description of languages in their phonemic aspect. *Language* 29:34–46.
- Chomsky, N. (1975). *The Logical Structure of Linguistic Theory*. New York: Plenum.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, N., and M. Halle. (1968). *The Sound Pattern of English*. New York: Harper & Row.
- Churma, D. (1988). In defense of morpheme structure rules: evidence from vowel harmony. *BLS* 14:23–34.
- Clements, G. N. (1985). The geometry of phonological features. *Phonology Yearbook*. 2:225–252.
- Clements, G. N. (1987). Towards a substantive theory of feature specification. *Proceedings of NELS* 18:79–93.
- Clements, G. N. (1985). Akan vowel harmony: a non-linear analysis. In D. Goyvaerts, ed. *African linguistics: essays in memory of M.W.K. Semikenke*. Amsterdam: J. Benjamins.
- Clements, G. N., and E. Sezer. (1982). Vowel and consonant disharmony in Turkish. In H. van der Hulst and N. Smith, eds. *The Structure of Phonological Representations, Part II*. Dordrecht: Foris.
- Clements, G. N., and R. Sonaiya. (1989). Underlying feature specification in Yoruba. Paper presented to the 20th Annual Conference on African Linguistics, University of Illinois, Urbana, Illinois.
- Coleman, J. (1991). *Phonological Representations — their names, forms and powers*. PhD dissertation: University of York.
- Coleman, J., and J. Local. (1989). The 'no crossing constraint' in autosegmental phonology. *York Papers in Linguistics* 14:169–219.
- Cook, T. (1986). The chamelonic vowel in the harmonizing prefixes of Efik. In K. Bogers, H. van der Hulst, and M. Mous (eds) *The Phonological Representation of Suprasegmentals* (Dordrecht: Foris) pp. 209–232.
- Davey, B. A., and H. A. Priestley. (1990). *Introduction to Lattices and Order*. Cambridge: C.U.P.
- Etherington, D., and R. Reiter. (1983). On inheritance hierarchies with exceptions. *Proceedings of the National (U.S.) Conference on Artificial Intelligence (AAAI-83)*, pp. 104–108.
- Fahlman, S. (1979). *NETL: A system for Representing and Using Real-World Knowledge*. Cambridge, Mass.: MIT Press.
- Firth, J. R. (1957). A synopsis of linguistic theory, 1930–55. In *Studies in Linguistic Analysis* [Special Volume of the Philological Society]. Oxford: Blackwell.
- Ganter, B., K. Rindfrey, and M. Skorsky. (1986). Software for formal concept analysis. In Gaul and Schader (1986).
- Gaul, W., and M. Schader, eds. (1986). *Classification as a Tool of Research*. North-Holland: Elsevier Science Publishers.
- Gazdar, G. ed. (1990). *Papers in DATR*. Dept. Cognitive Science, University of Sussex.

- Gazdar, G., E. Klein, G. Pullum and I. Sag. (1985). *Generalized Phrase Structure Grammar*. Oxford: Blackwell.
- Gibbon, D. (1990). Underspecification in phonology. In Gazdar (1990).
- Goldsmith, J. (1976). Autosegmental Phonology. PhD dissertation: MIT. [Published New York: Garland, 1979].
- Goldsmith, J. (1987). Vowel systems. In *23rd Annual Regional Meeting of the Chicago Linguistic Society. Part II: Parasession on Autosegmental and Metrical Phonology*, pp. 116–133.
- Goldsmith, J. (1989). The aims of autosegmental phonology. In D. Dinnsen, ed. *Current Approaches to Phonological Theory*. Bloomington: Indiana University Press.
- Goldsmith, J. (1989). Licensing, inalterability and harmonic rule application. *Proceedings of the Chicago Linguistic Society* 25:145–156.
- Gordon, P., J. Eberhardt, and J. Rueckl. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology* 25:1–42.
- Halle, M. (1959). *The Sound Pattern of Russian: a linguistic and acoustical investigation*. 'S-Gravenhage: Mouton & Co.
- Halliday, M. (1961). Categories of the theory of grammar. *Word* 17:241–292.
- Hayes, B. (1986). Inalterability in CV phonology. *Language* 62:321–351.
- Henrici, A. (1981). Some notes on the systemic generation of a paradigm of the English clause. In M.A.K. Halliday and J. Martin, eds. *Readings in Systemic Linguistics*. London: Batsford
- Hess, S. (1988). Acoustic characteristics of the vowel harmony feature and vowel raising in Akan. *UCLA Working Papers in Phonetics* 25:58–72.
- Hockett, C. (1987). *Refurbishing our Foundations: elementary linguistics from an advanced point of view*. Amsterdam: J. Benjamins.
- Horty, J., R. Thomason, and D. Touretzky. (1987). A skeptical theory of inheritance in nonmonotonic semantic nets. *Proceedings of the National (U.S.) Conference on Artificial Intelligence (AAAI-87)*, pp. 322–325.
- Hualde, J. (1991). Unspecified and unmarked vowels. *Linguistic Inquiry* 22:205–209.
- Huttenlocher, D. and V. Zue. (1983). *Proceedings of the National (U.S.) Conference on Artificial Intelligence (AAAI-83)*, pp. 172–176.
- Itô, J. (1986). *Syllable Theory in Prosodic Phonology*. PhD dissertation: University of Massachusetts. [Published New York: Garland, 1988].
- Itô, J., and R. Mester. (1989). Feature predictability and underspecification: palatal prosody in Japanese mimetics. *Language*. 65:258–293.
- Kahn, D. (1976). *Syllable-based Generalizations in English Phonology*. PhD dissertation: MIT.
- Kenstowicz, M., and C. Kisseberth. (1979). *Generative Phonology: Description and theory*. San Diego: Academic Press.
- Kiparsky, P. (1965). *Phonological Change*. PhD dissertation: MIT.

- Kiparsky, P. (1982). Lexical morphology and phonology. In I.-S. Yang, ed. *Linguistics in the Morning Calm*. Seoul: Hanshin.
- Kiparsky, P. (1985). Some consequences of lexical phonology. *Phonology Yearbook* 2:82–138.
- Kisseberth, C. (1972). Cyclical rules in Klamath phonology. *Linguistic Inquiry* 3:3–33.
- Lass, R. (1976). *English Phonology and Phonological Theory*. Cambridge: C.U.P.
- Lightner, T. (1973). Against morpheme structure rules and other things. In *Issues in Phonological Theory: Proceedings of the Urbana conference on phonology*. The Hague: Mouton.
- Lindau-Webb, M. (1988). Tongue mechanisms in Akan and Luo. *UCLA Working Papers in Phonetics* 25:46–57.
- Luce, P., Feustal, T., and Pisoni, D. (1983). Capacity demands in short-term memory for synthetic and natural speech. *Human Factors* 25:17–32.
- Luksch, P., M. Skorsky, and R. Wille. (1986). On drawing concept lattices with a computer. In Gaul and Schader (1986).
- Macdonell, A. (1927). *Sanskrit Grammar*. Oxford: O.U.P.
- Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: C.U.P.
- McCawley, J. (1968). Concerning the base component of a transformational grammar. *Foundations of Language*. 4:243–269.
- McCawley, J. (1970). Meaningful noises. *Times Literary Supplement* 3569:817–18.
- Mellish, C. (1988). Implementing systemic classification by unification. *Computational Linguistics* 14:40–51
- Mohanan, K. P. (1991). On the bases of radical underspecification. *Natural Language & Linguistic Theory*. 9:285–326.
- Moshier, M., and W. Rounds. (1987). A logic for partially specified data structures. In *ACM Symposium on the Principles of Programming Languages*. Munich, Germany.
- Nutter, J. (1983). Default reasoning using monotonic logic: a modest proposal. *Proceedings of the National (U.S.) Conference on Artificial Intelligence (AAAI-83)*, pp. 297–300.
- Omamor, A. (1988). Okpe and Uvwie: a case of vowel harmony galore. *Journal of West African Languages* 18:47–64.
- Paradis, C., and J-F. Prunet, eds. (1991). *Phonetics and Phonology 2: The Special Status of Coronals*. San Diego: Academic Press.
- Pierrehumbert, J. (1992) Dissimilarity in the Arabic Verbal Roots. In *Proceedings of NELS* 23.
- Pollard, C., and I. Sag. (1987). *Information-based Syntax and Semantics*. Stanford: CSLI.
- Prince, A. (1984). Phonology with tiers. In M. Aronoff and R. Oehrle, eds. *Language sound structure: studies in phonology*. Cambridge, Mass.: MIT Press.
- Pulleyblank, D. (1983). *Tone in Lexical Phonology*. PhD dissertation: MIT. [Published Dordrecht: Reidel, 1986].

- Pulleyblank, D. (1986). Underspecification and low vowel harmony in Okpe. *Studies in African Linguistics* 17:119–153.
- Pulleyblank, D. (1988). Vocalic underspecification in Yoruba. *Linguistic Inquiry* 19:233–270.
- Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence* 13:81–132.
- Roberts, R., and I. Goldstein. (1977). The FRL Manual. AI Memo No. 409, MIT Artificial Intelligence Laboratory, Cambridge, Mass.
- Robins, R. (1957). Aspects of Prosodic Analysis. *Proceedings of the University of Durham Philosophical Society*. Series B. I (Arts), 1:1–12.
- Sagey, E. (1986). *The Representation of Features and Relations in Non-linear phonology*. PhD dissertation: MIT.
- Sandewall, E. (1986). Nonmonotonic inference rules for multiple inheritance with exceptions. *Proceedings of the IEEE* 74:1345–1353.
- Shephard, R. N. (1984). Ecological constraints on internal representation: resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review* 91:417–447.
- Shephard, R. N. (1987a). Evolution of a mesh between principles of the mind and regularities of the world. In J. Dupré, ed. *The Latest on the Best: Essays on Evolution and Optimality*. Cambridge, Mass.: MIT press.
- Shephard, R. N. (1987b). Toward a universal law of generalization for psychological science. *Science* 237:1317–1323.
- Shephard, R. N. (1989). Internal representation of universal regularities: a challenge for connectionism. In L. Nadel, L. A. Cooper, P. Culicover, and M. Harnish, eds. *Neural Connections, Mental Computation*. Cambridge, Mass.: MIT press.
- Shibatani, M. (1973). The role of surface phonetic constraints in generative phonology. *Language* 49:87–106.
- Shibatani, M. and J. Crothers. (1974). On the status of blank features in phonology. *Glossa* 8:261–270.
- Shieber, S. (1986). *An Introduction to Unification-based Approaches to Grammar*. Stanford: CSLI.
- Shieber, S., H. Uszkoreit, J. Robinson, and M. Tyson. (1983). The formalism and implementation of PATR-II. In *Research on Interactive Acquisition and Use of Knowledge*. Menlo Park, Calif.: Artificial Intelligence Center, SRI International.
- Shipman, D., and V. Zue. (1982). Properties of large lexicons: implications for advanced isolated word recognition systems. Conference record, IEEE International Conference on Speech Acoustics and Signal Processing. Paris, France, pp.546–549.
- Sprigg, R. (1965). Prosodic analysis and Burmese syllable-initial features. *Anthropological Linguistics* 7:59–81.
- Sproat, R. (1992). *Morphology and Computation*. Cambridge, Mass.: MIT Press.
- Stahl, J., and R. Wille. (1986). Preconcepts and set representation of contexts. In Gaul and Schader (1986).



- Stanley, R. (1967). Redundancy rules in phonology. *Language* 43:393–436.
- Steriade, D. (1987). Redundant values. In *23rd Annual Regional Meeting of the Chicago Linguistic Society. Part II: Parasession on Autosegmental and Metrical Phonology* pp. 339–362
- Steriade, D. (1988). Greek accent: a case for preserving structure. *Linguistic Inquiry*. 19:271–314.
- Stewart, J. (1971). Niger-Congo, Kwa. In T. Sebeok, ed. *Current Trends in Linguistics* [Volume 7: Linguistics in Sub-Saharan Africa]. The Hague: Mouton.
- Stockwell, R., D. Bowen, and J. Martin. (1965). *The Grammatical Structures of English and Spanish*. Chicago: University of Chicago Press.
- Teder, H. (1990). *Hearing Instruments* 41, p.32.
- Touretzky, D., (1984). Implicit ordering of defaults in inheritance systems. *Proceedings of the National (U.S.) Conference on Artificial Intelligence (AAAI-84)*, pp. 322–325.
- Touretzky, D. (1986). *The Mathematics of Inheritance Systems*. Los Altos, Calif.: Morgan Kaufman Publishers.
- Touretzky, D., J. Horty, and R. Thomason. (1987). A clash of intuitions: the current state of nonmonotonic multiple inheritance systems. *IJCAI* 87:476–482.
- van Riemsdijk, H. and E. Williams. (1986). *Introduction to the Theory of Grammar*. Cambridge, Mass.: MIT Press.
- Wardrip-Fruin, C. (1985). The effect of signal degradation on the status of cues to voicing in utterance-final stop consonants. *Journal of the Acoustic Society of America* 77:1907–1912.
- Waterson, N. (1981). A tentative developmental model of phonological representation. In T. Myers, J. Laver, and J. Anderson, eds. *The Cognitive Representation of Speech*. Amsterdam: North-Holland. [Reprinted in Waterson (1987)].
- Waterson, N. (1987). *Prosodic Phonology: the theory and its application to language acquisition and speech processing*. Newcastle-upon-Tyne: Grevatt & Grevatt.
- Wille, R. (1984). Line diagrams of hierarchical concept systems. *International Classification* 11:77–85.
- Wille, R. (1989). Lattices in data analysis: how to draw them with a computer. In I. Rival, ed. *Algorithms and Order*. Kluwer.
- Winograd, T. (1972). *Understanding Natural Language*. San Diego: Academic Press.
- Yip, M. (1989). Feature geometry and cooccurrence restrictions. *Phonology* 6:349–374.
- Zwicky, A. (1965). *Topics in Sanskrit Phonology*. PhD dissertation: MIT.